2018

Elucidation Of Histone Modifications And Nucleosomal Structure Using Novel Mass Spectrometry Approaches

Kelly Karch
University of Pennsylvania, karchk1@gmail.com

Elucidation Of Histone Modifications And Nucleosomal Structure Using Novel Mass Spectrometry Approaches

Abstract
The fundamental repeating unit of chromatin is the nucleosome, composed of 147 base pairs of DNA wrapped around a histone protein octamer containing two copies of H2A, H2B, H3, and H4. Histone proteins are involved in many critical nuclear processes including transcription and maintenance of chromatin structure. Histone function is mediated by a dynamic and extensive array of post-translational modifications (PTMs). Mass spectrometry (MS) has emerged as a leading tool to study these complex histone PTM profiles. Generally, MS experiments utilize data dependent acquisition (DDA) methods on high-resolution MS instruments because they can more readily distinguish PTMs with small mass differences. I demonstrate here that low-resolution instruments are capable of this analysis with data dependent acquisition (DDA) and data independent acquisition (DIA) methods, thereby expanding the repertoire of instruments that can be used. However, DIA methods improve quantification of isobaric peptides compared to DIA and also allows for re-mining of data post-experiment. This dissertation also highlights work I have done to develop MS methods to identify and quantify ADP-ribosylation PTMs, which are critical for DNA damage repair pathways. We identified 30 ADP-ribosylation marks on histones, 20 of which are novel. We quantified 10 of these sites throughout a DNA damage and found that all of these sites increase in abundance over time, indicating that it is unlikely that specific sites are required for repair, but rather that ADP-ribosylation of the nucleosome surface in general is needed.

Histone function is also mediated through its structure and dynamic properties. Hydrogen-deuterium exchange (HDX) coupled to MS is a powerful technique to monitor these properties in solution. However, traditional HDX-MS studies on histone proteins were unable to monitor histone N-terminal tail domains, where a majority of PTM sites are located. Here, we demonstrate that by incorporating electron transfer dissociation (ETD) MS/MS methodology with middle-down and top-down MS, we are able to measure deuterium content of tail domains with near site-specific resolution for the first time. We find that all tails undergo decreased structural rigidity upon incorporation into the nucleosome, lending the first detailed experimentally-obtained insight into histone tail structure in solution.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biochemistry & Molecular Biophysics

First Advisor
Benjamin A. Garcia

Second Advisor
David Speicher

Keywords
ADP-ribosylation, Data Independent Acquisition, Histones, Hydrogen deuterium exchange, Mass spectrometry

Subject Categories
Analytical Chemistry | Biochemistry | Cell Biology

ELUCIDATION OF HISTONE MODIFICATIONS AND NUCLEOSOMAL STRUCTURE USING NOVEL MASS SPECTROMETRY APPROACHES

Kelly R. Karch

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

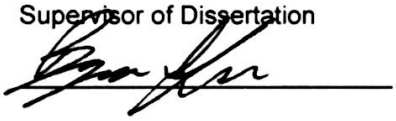
in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

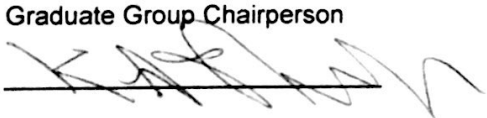
Supervisor of Dissertation



Benjamin A. Garcia

Presidential Professor of Biochemistry and Biophysics

Graduate Group Chairperson



Kim Sharp, Associate Professor of Biochemistry and Biophysics

Dissertation Committee

David Speicher, Caspar Wistar Professor in Computational and Systems Biology

Ronen Marmorstein, George W. Raiziss Professor

Ben E. Black, Associate Professor of Biochemistry and Biophysics

Kojo S. J. Elenitoba-Johnson, M.D., Peter C. Nowell, M.D. Professor

For my mother, father, and sister.

ACKNOWLEDGMENTS

Science is not a solitary pursuit- it takes many people to make a new discovery, and I have many people to thank for helping me reach my goals and making the journey fun and fulfilling.

I first have to thank Ben for being my mentor. I count myself extremely lucky to have such a supportive, understanding, and kind PI. Ben encouraged me to learn for myself but constantly made himself available to provide guidance. From him, I learned how to be critical and ask the right questions. He believed in my scientific abilities from the beginning, and he pushed me to do projects I never would have dreamed of doing five years ago. I would also like to thank Ben for being an incredibly supportive PI- over my time here, he has allowed me to TA twice, mentor high school, undergraduate, and graduate students, take a journalism class and a writing internship, knowing it would take some time away from lab but that it would help me explore career options and develop professionally. He leads by example and has created a friendly, collaborative, and fostering lab environment, and for that, I am extremely grateful.

To my committee members- Dr. David Speicher, Dr. Ronen Marmorstein, and Dr. Ben Black- thank you for your support, encouragement, and wise advice. Your guidance has made my science stronger, and your advice has been helpful both scientifically and professionally.

I also need to thank my collaborators. To Dr. Ben Black and his lab members- especially Levani Zandarashvili, Morgan Gerace, Allan Zong, Jennine Dawicki-McKenna, Sam Falk, and Nikolina Sekulic- thank you for teaching me everything I know about protein purification and HDX and letting me use your equipment, reagents, and expertise to drive my project forward. Your advice and suggestions have been really valuable, and I had so much fun learning from you all. Also, thank you to Jamie DeNizio for being the best lab rotation mentor I could imagine and teaching me your amazing organizational skills. You're a brilliant scientist, and I learned so much from you! To Dr. Walter Englander, Leland Mayne, and Zhongyuan Kan- thank you for your many hours helping me with data analysis and experimental setup. To my collaborators Dr. John Pascal

and Dr. Marie-France Langelier- thank you for your support and guidance. My project would not have been possible without your help and advice.

I would also like to thank Dr. Barry Zee, Dr. Rosie Molden, and Dr. Laura-Mae Britton, the graduate students who mentored me during my rotation and in my first year in the lab. Despite moving from Princeton to Philadelphia to help Ben start his lab here and finishing their dissertation work, Barry, Rosie, and Laura-Mae spent a lot of time helping me learn the ropes. Their passion and creativity instilled in me an even deeper love of science. To Laura-Mae, thank you for your unwavering support, both in the science world and in life. I would not have made it without you, and I'm so grateful for your lasting friendship. We will always be bay mates in spirit.

To my lab mates, past and present- what would I have done without you? During the highs and lows of my graduate school experience, you have been present, supportive, and all around amazing, intelligent, and kind people. Your support and friendship has not only made my experience possible, but you've also made it fun. From crossword puzzle lunches to happy hours to being distracted by cat pictures in lab to helping solve science problems, you all helped the Garcia lab win the "happiest lab on the floor" award. You are all such talented, creative scientists, and I'm lucky to have learned from each one of you. I still find it amazing that we're able to cram so many of us in such a small space and still manage to like each other. To Amanda- thank you for your fun-loving, spontaneous, and thoughtful spirit. You taught me to be grounded and not take things so seriously. I miss you and think about you every day. To Monster (Mariel Coradin)- I'm so happy we got to be bay mates for most my time in the lab- your genuine, fun-loving, curious personality has made lab so much fun. You have the best sense of humor and I will miss playing pranks on people with you. To the rest of my bay mates- little Mariel, Laura, Peder- thank you for being awesome people. You've really made it fun to show up to work everyday. To Bhanu, our lab mom, we would fall apart without you. Thank you for keeping us on our toes and always coming to lab with a positive attitude and a smile.

I've made a lot of amazing friends during my time in graduate school. To Mara Olenick, Enrique Lin Shiao, and Ari Kahn- you guys are awesome. Thank you for always being up for an

adventure and being great friends. To Sam Falk- thank you for your unwavering friendship and fun shenanigans. Your support always comes when its most needed, and you always know the right thing to say. I have a feeling we'll still be braiding our hair together when we're old. To James Townsend, Robert Magin, and Joe Jordan- your friendship has made my life better. I will miss our aimless wanderings around Chinatown and late-night trips to Midtown. To Joe Jordan- your passion has inspired me to make change at Penn and in the world. You are one of the coolest and most genuine people I ever met, and thank you for always listening and supporting me.

To my family- I owe you my all. To Mom, Dad, and Laura- thank you for setting such great examples in my life. From you, I've learned to be independent and strong. Thank you for encouraging me at every turn and helping me pursue my passion. Your unwavering support means the world to me, and I love you all. To Laura- you are my best friend and all-around coolest person on the planet. I'm so lucky to have you. Thank you for inspiring me every day and being there for me without fail. I don't know where I'd be without you (probably dressing like a hobo). To my mother- you are such an inspiration. You've built an amazing career and a strong family and make a mean bean soup. You always make me feel better when things get hard. To me, you are super woman and I love you with all my heart. To my dad- you've instilled a passion for learning in me from a young age. You've taught me to be curious and inquisitive and showed me how to challenge myself. You taught me to stand strong in who I am. You make the world a better place, and I'm so happy I get to be your daughter. Every day, I see a little more of each of you in myself, and I couldn't be happier for it.

ABSTRACT

ELUCIDATION OF HISTONE MODIFICATIONS AND NUCLEOSOMAL STRUCTURE USING NOVEL MASS SPECTROMETRY APPROACHES

Kelly R. Karch

Benjamin A. Garcia

The fundamental repeating unit of chromatin is the nucleosome, composed of 147 base pairs of DNA wrapped around a histone protein octamer containing two copies of H2A, H2B, H3, and H4. Histone proteins are involved in many critical nuclear processes including transcription and maintenance of chromatin structure. Histone function is mediated by a dynamic and extensive array of post-translational modifications (PTMs). Mass spectrometry (MS) has emerged as a leading tool to study these complex histone PTM profiles. Generally, MS experiments utilize data dependent acquisition (DDA) methods on high-resolution MS instruments because they can more readily distinguish PTMs with small mass differences. I demonstrate here that low-resolution instruments are capable of this analysis with data dependent acquisition (DDA) and data independent acquisition (DIA) methods, thereby expanding the repertoire of instruments that can be used. However, DIA methods improve quantification of isobaric peptides compared to DIA and also allows for re-mining of data post-experiment. This dissertation also highlights work I have done to develop MS methods to identify and quantify ADP-ribosylation PTMs, which are critical for DNA damage repair pathways. We identified 30 ADP-ribosylation marks on histones, 20 of which are novel. We quantified 10 of these sites throughout a DNA damage and found that all of these sites increase in abundance over time, indicating that it is unlikely that specific sites are required for repair, but rather that ADP-ribosylation of the nucleosome surface in general is needed.

Histone function is also mediated through its structure and dynamic properties. Hydrogen-deuterium exchange (HDX) coupled to MS is a powerful technique to monitor these properties in solution. However, traditional HDX-MS studies on histone proteins were unable to

monitor histone N-terminal tail domains, where a majority of PTM sites are located. Here, we demonstrate that by incorporating electron transfer dissociation (ETD) MS/MS methodology with middle-down and top-down MS, we are able to measure deuterium content of tail domains with near site-specific resolution for the first time. We find that all tails undergo decreased structural rigidity upon incorporation into the nucleosome, lending the first detailed experimentally-obtained insight into histone tail structure in solution.

Table of Contents

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: Analyzing chromatin composition with mass spectrometry

In Eukaryotic cells, DNA is condensed into the nucleus in a highly organized fashion. This condensation is quite challenging from a molecular viewpoint given that Eukaryotic DNA is typically dozens of centimeters long and must fit into the micrometer-sized nucleus (Teif and Bohinc, 2011). Furthermore, DNA is a highly negatively charged molecule and so condensation into the nucleus poses a major electrostatic challenge (Carrivain et al., 2012).

DNA condensation in the Eukaryotic nucleus is achieved through the action of DNA-binding proteins (Carrivain et al., 2012). The full complement of DNA and its associated proteins is called chromatin. Chromatin is a highly dynamic and organized protein-DNA complex that is able to respond to cellular environment through structural or compositional changes. Some of these changes can be very dramatic; for example, during cellular division, chromatin undergoes a massive structural rearrangement to form highly condensed chromosomes. However, even smaller processes, such as transcription of a single gene, often require structural rearrangements and recruitment of specialized proteins (Perdigoto, 2017). Indeed, research on chromatin structure and composition and its link to many nuclear processes is a highly active area of research.

Chromatin structure is tightly linked to function. It has been known for decades that tightly condensed chromatin leads to repression of genes located in that region. These highly condensed regions of chromatin are called heterochromatin. Conversely, looser, more open chromatin structure is linked to gene transcription, likely because transcription machinery is able to physically access the DNA. Open, transcriptionally active chromatin is referred to as euchromatin. Chromatin binding proteins can alter chromatin structure and consequently transcription and other nuclear processes (Even-Faitelson et al., 2016).

The smallest repeating unit of chromatin is the nucleosome, which is composed of 147 base pairs of DNA wrapped around an octamer of histone proteins containing two copies of each

core histone- H2A, H2B, H3, and H4 (Figure 1.1) (Luger et al., 1997). Histone proteins are among the most basic proteins in the cell, containing a large number of lysine and arginine residues. These residues impart a positive charge on the histone proteins, which aids in the interaction with negatively charged DNA and reduces charge repulsion between nearby nucleosomes. Within the nucleosome, histones interact with each other through a hydrophobic, globular histone fold domain located in the C-terminal portion of the protein. Each histone also contains a flexible tail domain at the N-terminus that protrudes out from the nucleosomal surface (Arents and Moudrianakis, 1995).

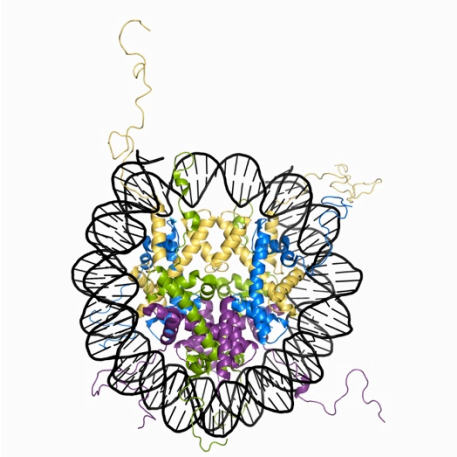


Figure 1.1. Crystal structure of the nucleosome. The nucleosome is composed of approximately 147 base pairs of DNA (black) and 2 copies of each core histone: H3 (yellow), H4 (blue), H2A (green), and H2B (purple). The tail domains, which protrude from the nucleosomal surface are modeled in (Davey et al., 2002). PDB: 1kx5.

Due to their intimate association with DNA, histone proteins are involved in nearly every nuclear process involving DNA, including gene transcription, cell division, and maintenance of higher order chromatin structure (Jenuwein and Allis, 2001). However, as non-enzymatic proteins, histone function is mainly mediated through protein interactions and a vast and dynamic

array of post-translational modifications (PTMs). A vast majority of these PTMs occur on the N-terminal tail domains of the histone proteins, especially histones H3 and H4 (Bannister and Kouzarides, 2011). As such, there is a large interest in the field to understand the PTM composition of histone proteins and how these PTMs affect histone function and nuclear processes.

Histone tails, in addition to being the major sites of post-translational modification, are critical for forming and maintaining higher order chromatin structure, such as the 30-nm chromatin fiber. In solutions containing mono- and divalent ions at physiological concentrations, nucleosome arrays will spontaneously fold into structures similar to the 30-nm fiber and can reversibly self-associate into higher order tertiary structures, demonstrating that the histone proteins alone can define higher order chromatin structures (Hansen, 2002; Woodcock and Ghosh, 2010). Linker histone H1, although it is not required for this process, promotes chromatin condensation and the formation of secondary and tertiary structures (Happel and Doenecke, 2009). The core histone tail domains have been shown to be critical for chromatin compaction, albeit to varying degrees. Studies in which histone tails are removed indicate that the H4 tail provides the largest contribution to formation and maintenance of higher order structure, while the H3 tail contributes slightly less to secondary and tertiary structure formation. H2A and H2B N-terminal tail domains are not required for chromatin condensation, but still contribute to the formation of these structures (Dorigo et al., 2003; Gordon et al., 2005; Tse and Hansen, 1997). The H4 tail has been shown to bind to an acidic patch on the surface of H2A and H2B on an adjacent nucleosome, an interaction which is likely important for chromatin folding (Dorigo et al., 2003; Kalashnikova et al., 2013; Luger et al., 1997). However, the structures and interactions of histone tail domains and the mechanisms by which they participate in chromatin folding pathways have not been well characterized.

Histone proteins are among the most post-translationally modified proteins present in Eukaryotic cells, and the suite of PTMs that can occur on histones is vast and diverse. PTMs can

range in size from a single methyl mark to the addition of entire proteins, such as ubiquitin, to ADP-ribose polymers, which can contain up to hundreds of ADP-ribose units (Zhao and Garcia, 2015). To date, dozens of novel histone PTMs have been identified, and new ones are still being discovered today (Dai et al., 2014; Simithy et al., 2017; Tan et al., 2011). The roles of these PTMs in mediating nuclear functions are still being actively researched today.

Given the complexity of histone PTM profiles and the range of nuclear processes mediated by histones, it has been hypothesized that histone PTMs form a "histone code" in which specific PTMs and combinations of PTMs are responsible for mediating a specific physiological response by regulating chromatin structure or composition (Jenuwein and Allis, 2001). Indeed, many chromatin states contain defined histone PTM profiles, indicating that histone PTMs are at least partially responsible for regulating and defining chromatin function (Figure 1.2). Misregulation of histone PTMs have also been linked to many diseases, including cancer, highlighting their importance in maintaining nuclear integrity (Portela and Esteller, 2010).

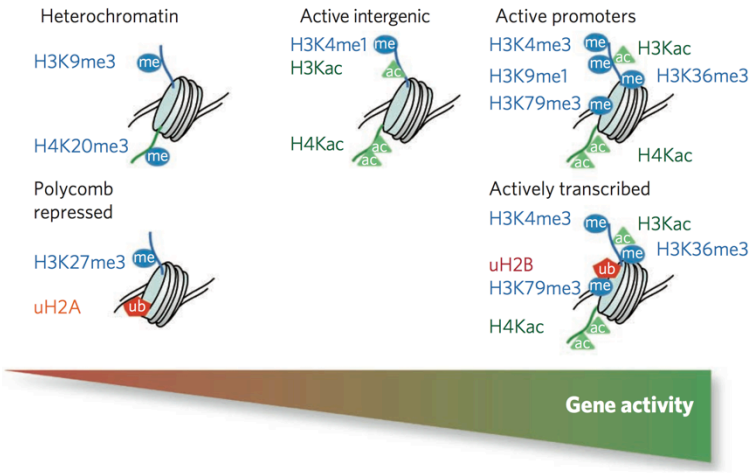


Figure 1.2. Chromatin states can be defined by histone PTM profiles. Figure taken from (Fierz and Muir, 2012).

Some PTMs regulate chromatin structure by directly altering the chemical environment. For example, acetylation neutralizes the positive charge of a lysine residue side chain, which can directly destabilize interaction with DNA and consequently cause chromatin to adopt a more open conformation (Bannister and Kouzarides, 2011). Furthermore, ubiquitinylation of H2B, a mark associated with transcriptional elongation, results in a decrease in chromatin compaction, likely due to the steric bulk of the ubiquitin moiety (Fierz et al., 2011). The structural basis of PTM function for most marks, however, is still largely unexplored.

Many PTMs exert their function by recruiting or blocking chromatin modifying proteins to the modification site. Proteins that bind to histone PTMs are generally referred to as "readers", while proteins that remove or add the PTM are referred to as "erasers" or "writers," respectively (Fierz and Muir, 2012). For example, tri-methylated H3K9 recruits heterochromatin binding protein 1 (HP1) to promote formation of condensed, transcriptionally silent chromatin. However, if the adjacent residue, H3S10, is phosphorylated, HP1 is physically blocked from interaction, allowing chromatin to adopt a more open form (Fischle et al., 2005). Therefore, in order to fully understand the biological state of a specific region of chromatin, elucidating the complete histone PTM profile is desirable.

1.1. Traditional methods to study histone PTMs

Histone biology has been an active area of research for decades. Vincent Allfrey and colleagues discovered the first histone PTMs, acetylation and methylation, over 50 years ago (Allfrey et al., 1964). The researchers identified these PTMs by incubating calf thymus nuclei with ^{14}C-labeled sodium acetate and monitoring the incorporation of radiolabeled carbon into histone proteins.

Since then, new methods have emerged to address histone PTM content and localization in chromatin. Many of these methods are antibody-based, such as Western blot, chromatin

immunoprecipitation (ChIP), and ChIP-seq (Britton et al., 2011). In Western blot techniques, a sample is separated by gel electrophoresis and transferred to a membrane where an antibody can be used to determine if a PTM of interest is present. ChIP-qPCR and ChIP-seq methods can be used to determine where a PTM of interest is located in the genome, usually to determine which genes are regulated by the modification. In ChIP methods, proteins are crosslinked to DNA, which is subsequently sheared by sonication. Then, a protein or PTM of interest is purified using an antibody. The sample can then be un-crosslinked and the purified DNA can then either be sequenced and mapped to the genome (ChIP-seq), or primers can be used to detect and quantify the presence of a specific DNA sequence of interest (ChIP-qPCR).

However, there are several disadvantages associated with the use of antibody-based methods such as these for studying PTMs. Firstly, development of modification-specific antibodies is difficult, costly, and challenging to validate. One major concern is cross-reaction with similar modifications (i.e. mono- versus di-methylation) or the same modification located in a different region of the same histone or different histone protein (i.e. H3K9me3 versus H4K20me3). Epitope occlusion is also common, where the epitope of an antibody is blocked by the presence of nearby PTMs. This issue is particularly relevant to histone proteins where many PTM sites are located in close proximity. Another challenge is that a priori knowledge of PTMs is required for generation of antibodies and therefore cannot be used to discover novel PTMs. Furthermore, antibody-based methods generally probe a single PTM at a time and therefore cannot provide a global view of histone PTM profiles. Each of these drawbacks makes it incredibly challenging to accurately and quantitatively address histone PTM content using antibodies (Fuchs et al., 2011; Rothbart et al., 2015).

Mass spectrometry (MS)-based methods, on the other hand, offer a quantitative and unbiased approach to study histone modifications at a global level. MS cannot, however, localize a PTM of interest to specific region of the genome. One of the greatest advantages of MS is the potential to identify novel PTMs and measure the co-occurrence of modifications on the same

peptide or protein. As such, MS has emerged as one of the most powerful tools for histone modification analysis.

1.2. Mass-spectrometry based methods to study histone PTMs

1.2.1. Principles of Liquid Chromatography and Mass Spectrometry

A schematic of the general layout of the mass spectrometry experiments performed in this work is shown in Figure 1.3. There are many different types of mass spectrometers, but only the type used in the experiments described in this thesis will be discussed. A mass spectrometer is composed of three parts: an ion source, mass analyzer, and detector.

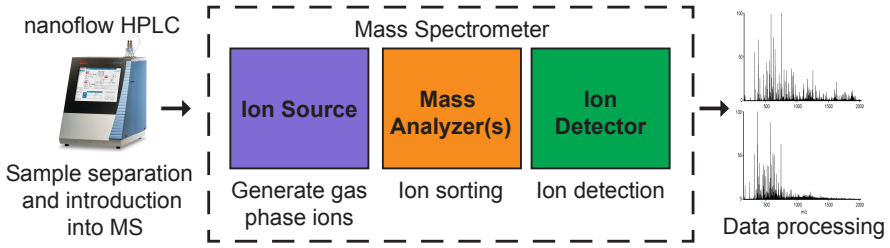


Figure 1.3. General schematic of mass spectrometry experiment.

In order for an analyte to be analyzed by MS, it must be charged and enter the gas phase. The ion source is responsible for this process. There are many different types of ionization, but the type used in this work is called electrospray ionization (ESI, Figure 1.4). Proteins and peptides carry a net positive charge, and this can be ensured by the addition of acids to the buffer. In ESI, samples are sprayed into a fine mist that then enters the mass spectrometer. The entrance to the mass spectrometer, called a capillary, is heated to

approximately 200-350˚C. The heat from this capillary allows the droplets of sample to begin evaporation. At a certain point, the Columbic repulsion between positively charged analyte molecules will be greater than the surface tension of the droplet, and the analyte ions will enter the gas phase. A strong electric field is generated by applying a positive voltage to the sample outlet and a negative voltage to the mass spectrometer inlet. This electric field facilitates ionization and directs the analyte ions into the mass spectrometer (Banerjee and Mazumdar, 2012; Fenn et al., 1989).

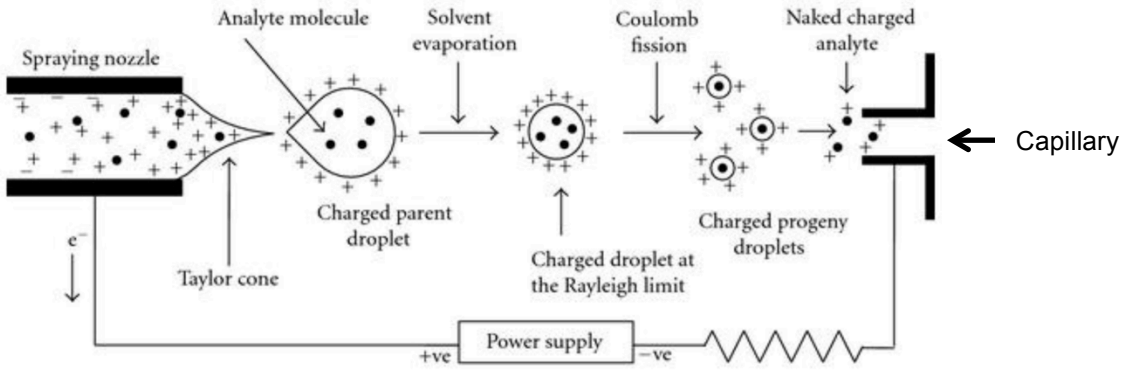


Figure 1.4. Schematic of electrospray ionization. Figure modified from (Banerjee and Mazumdar, 2012).

The mass analyzer separates analyte ions based on their mass to charge ratio (m/z). There are several types of mass analyzers, but the ones used in this study are the quadrupole and linear ion trap mass analyzers. The quadrupole is composed of four parallel cylindrical rods. A radio frequency (RF) voltage is applied to the rods and alternated to create an oscillating electric field. Analyte ions can travel down the quadrupole in the center of the rods. However, only ions in a certain m/z range will be stable in the given voltage ratio between the rods, and all others will be ejected through the spaces between the rods and therefore not detected/measured (Figure 1.5A). Therefore, quadrupoles can be used to select ions with a specific m/z or ranges of

m/z. Ions that are filtered by a quadrupole can be directed to the detector or to another mass analyzer (Gilany and Luc, 2010). A linear ion trap mass analyzer works similarly, except that ions are first trapped and then ejected sequentially. The ion trap is composed of four electrode plates with alternating voltages that creates an electric field (Figure 1.5B). The given ratio of voltages between the plates causes ions of resonant m/z to be "trapped," orbiting in the space between the plates, while ions of non-resonant m/z are ejected from the trap. The RF can be increased over time, allowing smaller m/z ions to be less stabilized and ejected (March, 1997).

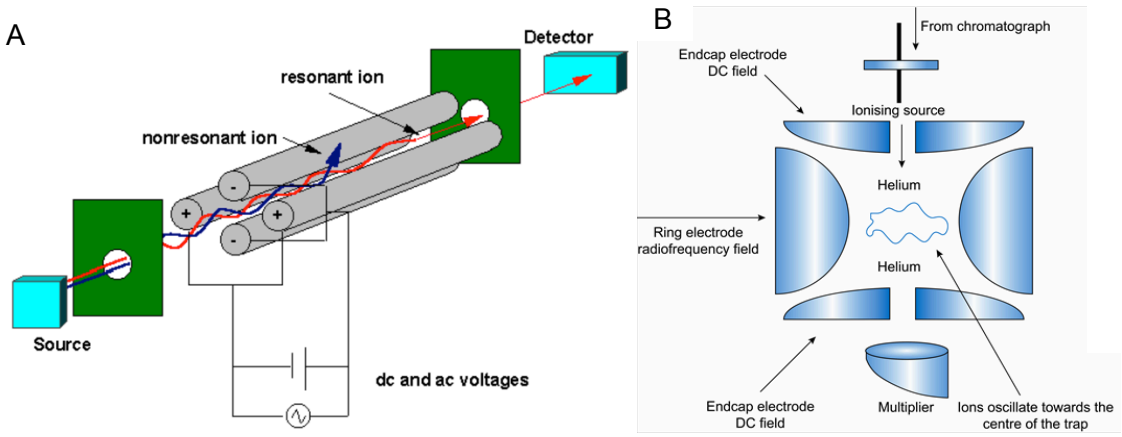


Figure 1.5. Schematic of quadrupole and ion trap mass analyzers. (A) Principles of quadrupole mass analyzer. Figure taken from (Gilany and Luc, 2010). (B) Principles of the ion trap mass analyzer. Figure taken from www.medicinescomplete.com.

The m/z values of ions are measured by the detector. An electron multiplier is a type of detector that is generally coupled to ion traps. As ions are sequentially ejected by the ion trap as the RF field increases, they will hit the electron multiplier, with can then record the signal. Orbitraps are a different type of detector that can measure the m/z of many ions at once (Hu et al., 2005). An Orbitrap consists of two electrodes, an inner, spindle-like electrode surrounded by outer electrode (Figure 1.6). An electric field is generated between the electrodes. A packet of

ions can be trapped in a different region of the mass spectrometer (called a C-trap) and injected into the Orbitrap. The electric field then causes the ions to oscilate around the inner electrode, and the rate at which they oscillate is dependent on the m/z. The current of the ions is measured and a Fourier Transform is used to determine the component m/z values (Hu et al., 2005; Makarov, 2000).

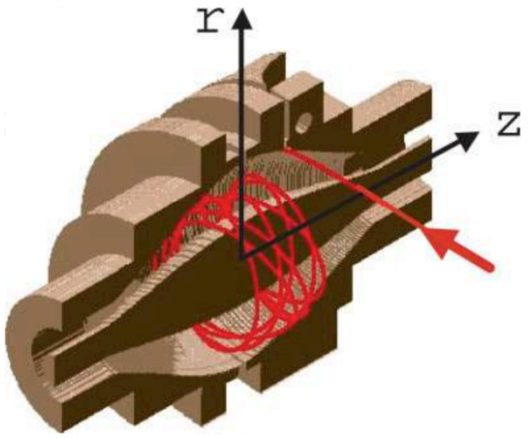


Figure 1.6. Schematic of the Orbitrap detector. The schematic shows an Orbitrap cut in half lengthwise. Ions are shown in red and are injected into the Obritrap as indicated by the arrow. Ions spin around the inner electrode along the r axis and oscillate back and forth according to their m/z values along the inner electrode in the z direction. Figure taken from (Hu et al., 2005).

Two different scan types are used in this thesis: full mass spectrum (alias: MS1), and a tandem mass spectrum (alias: MS2 or MS/MS). In an MS1 scan, the m/z values of all ions in a given range are measured and recorded, along with their relative intensities. In an MS2 scan, a single ion or groups of ions of interest are isolated, fragmented, and the m/z values of the resulting fragment ions are recorded. MS2 scans are critical for determining the amino acid

sequence of a peptide and can also be used to identify PTMs (Aebersold and Mann, 2003; Steen and Mann, 2004).

Usually, for peptide sequencing and PTM identification, a single ion of interest is selected for fragmentation. Generally speaking, a peptide will only break in one location on the backbone, generating two fragment ions, an N-terminal ion, and a C-terminal ion. The MS/MS spectrum can be used to identify the peptide based on the m/z values of the fragment ions. For example, if a peptide with the sequence KELLY was selected, there would be fragment ions corresponding to K, KE, KEL, KELL, Y, LY, LLY, ELLY. Therefore, the connectivity of the amino acid sequence can be determined. If there was a PTM located on a residue, for example, a hydroxamic acid moiety on E, then each fragment ion containing E would also exhibit a mass shift corresponding to the mass of hydroxamic acid. Thus, the PTM can be localized. Generally, search software is used to identify these fragments. The user imports a database containing all of the sequences of the potential proteins and PTMs in the sample to define the search. Then, many search softwares operate by generating a theoretical MS and MS/MS spectrum for each potential peptide based on the database, and tries to match the experimental spectra to the theoretical. Usually, a score is given based on how well the spectra match (Steen and Mann, 2004).

There are several different types of fragmentation methods. Here, I will discuss two types: collision-based and electron transfer-based fragmentation methods. In collision-based methods, such as collision induced dissociation (CID) and higher energy collisional dissociation (HCD), analyte ions are accelerated by an electrical potential to reach a high kinetic energy. Then, the ions are introduced to a chamber containing neutral gas molecules (such as helium, nitrogen, or argon) where they collide with these molecules. The collision converts some of the kinetic energy into internal energy, causing the ion to break in a single position on the backbone at the amide bond. The resulting fragment ions are called b ions (N-terminal) and y ions (C-terminal) (Mitchell Wells and McLuckey, 2005). CID and HCD fragmentation are by far the most commonly used methods for peptide dissociation, however these methods do have several notable drawbacks: (1) backbone cleavage is non-random, with cleavage occurring preferentially

at basic residues which can lead to incomplete sequence coverage, and (2) labile PTMs, such as phosphorylation, can be lost or preferentially fragmented over the backbone due to the high energy imparted to the molecule during the fragmentation process.

Electron transfer-based fragmentation methods, such as electron transfer dissociation (ETD) and electron capture dissociation (ECD), promote cleavage at the N-Cα bond of the peptide or protein backbone through the transfer of an electron to the analyte ion. In ECD methods, low energy electrons are directly transferred to the analyte in the ion cyclotron resonance cell of a Fourier transform mass spectrometer. ECD spectra are obtained by averaging data over many scans, and are therefore often not compatible with the liquid chromatography timescale. In ETD, on the other hand, an anion carrier reagent (such as fluoranthene), transfers the electron. ETD fragmentation can be implemented in a variety of instruments and is therefore more versatile than ECD (Syka et al., 2004; Udeshi et al., 2007; Wiesner et al., 2008). ETD is more efficient than ECD and therefore does not require the averaging of many scans. The resulting fragment ions of ETD and ECD fragmentation are called c ions (N-terminal) and z ions (C-terminal). ETD has several advantages over CID and HCD, including that it is not biased towards specific amino acids, thereby allowing for more even backbone cleavage and is also better suited to retain labile PTMs (Udeshi et al., 2007; Wiesner et al., 2008). One drawback, however, is that ETD and ECD are better suited for more highly charged analyte ions and therefore do not perform as well as CID or HCD on lower charged ions, for example +2 and +3 charged.

1.2.2. Different Types of Mass Spectrometry Experiments for Histone Analysis

Mass spectrometry (MS) is a diverse and highly adaptable technology. As such, there are many MS approaches to study proteins, namely bottom-up (BU) MS, middle-down (MD) MS, and top-down (TD) MS (Figure 1.7).

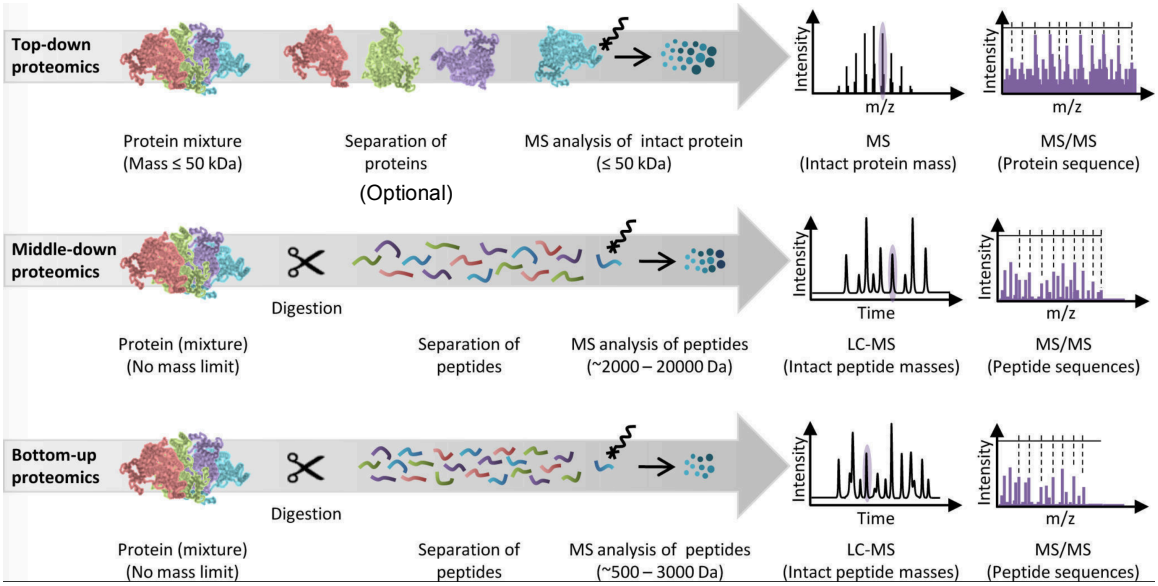


Figure 1.7. Workflow for bottom-up, middle-down, and top-down mass spectrometry experiments. Example proteases are given for middle-down, but any protease that generates long peptides can be used. Figure taken from (Switzar et al., 2013).

In a BU-MS experiment, samples are digested, usually with trypsin, to generate peptides approximately 5-20 amino acids long. The peptides are then chromatographically separated and analyzed by MS for identification and quantification. Usually, reversed-phase high performance liquid chromatography (RP-HPLC) is used to separate peptides. RP-HPLC uses a hydrophobic stationary phase and polar mobile phase. An increasing proportion of an organic phase is used to elute peptides based on their hydrophobicities (Walther and Mann, 2010).

Trypsin, which cleaves after lysine and arginine residues, is the digestion enzyme of choice in BU-MS because it generates peptides of desirable length for most proteins, is relatively inexpensive, and is very efficient. The high specificity and efficiency of trypsin allows for a pool of highly reproducible peptides to be generated .

Trypsin digestion poses problems for histone analysis, however. Histone proteins contain a much greater ratio of lysine and arginine residues compared to other proteins and are therefore over-digested by trypsin, and the small peptides that result from digestion do not retain well on RP columns. Furthermore, many of these lysine and arginine residues are adjacent to each other, resulting in many missed-cleavage events as trypsin will one cleave after one of the adjacent residues at random. Furthermore, modified lysine and arginine residues prevent trypsin cleavage, resulting in more missed-cleavage events. Therefore, a single modification sites can be found on many different peptides, making accurate quantification very challenging (Garcia et al., 2007a).

These challenges can be overcome by derivitizing the histone sample prior to digestion. There are currently many different derivitaztion techniques for histone analysis, but the most common utilizes propionic anhydride (Garcia et al., 2007a; Sidoli et al., 2015). Propionic anhydride derivatization will add a propionyl group to any unmodified or mono-methylated lysine residue side chains as well as the N-terminal amino group of the protein. This procedure ensures that all lysine residues cannot be cleaved by trypsin, either due to endogenous modification or due to the addition of the propionyl moiety to the side chain. Therefore, trypsin will only cleave after arginine residues, essentially mimicking an Arg-C digestion but with the efficiency and reproducibility of a trypsin digestion. After digestion, another round of derivatization is completed to propionylate the newly exposed peptide N-terminal amino groups. The propionyl groups also impart a greater hydrophobicity to the peptide, allowing for better retention and separation on RP columns (Garcia et al., 2007a). Tryptic peptides usually occupy charge states +2 or +3, and therefore CID or HCD fragmentation is usually used for identification.

BU-MS is by far the most common method used today to study histone proteins. However, one major caveat of BU-MS analysis is that it cannot measure the co-occurance of PTMs located on different peptides. For example, results may indicate that H3K9me3 and H3K27me3 are present at a high abundance, but it cannot be determined if these marks tend to occur on the same histone molecule or on different molecules because they are not located on

the same peptide. Middle-down and top-down MS are used to study larger peptides or intact proteins, respectively, and are therefore more useful for deciphering the histone combinatorial code. However, these approaches are more challenging than BU-MS, both in terms of experimentation and analysis.

Middle-down MS involves the generation and analysis of larger peptides (generally >20 amino acids). Therefore, proteases other than trypsin must be used. It is ideal to use a protease that cleaves after a single residue to reduce the amount of missed cleavage events. Endoproteinase AspN, which cleaves N-terminal to aspartate residues, can be used to generate the H4 1-24 peptide, which contains most of the known PTM sites on H4. Endoproteinase GluC, which cleaves C-terminal to glutamate residues, can be used to generate the H3 1-50 peptide, which contains many but not all of the known PTM sites on H3.

Middle-down analysis of histone proteins generally does not utilize RP chromatography because the hydrophilic nature of the histone tails results in poor binding to the hydrophobic stationary phase, and more importantly, poor separation of the histone tails. Weak cation exchange (WCX) hydrophilic interaction liquid chromatography (HILIC) has been used to achieve better separation of histone tails. WCX-HILIC employs a hydrophilic stationary phase coupled to an organic mobile phase, and can be used online or offline (Garcia et al., 2007b; Pesavento et al., 2008; Young et al., 2009).

Top-down MS involves the analysis of intact proteins and therefore complete connectivity of histone PTMs can be retained and analyzed. Although chromatography can be used for separation of intact proteins, most studies to date use direct infusion as it is difficult to achieve separation of differently modified intact histone proteins (Patrie, 2016).

One major caveat of middle-down and top-down MS is that the sensitivity is lower. Longer peptides and intact proteins have a larger number of potential modification states and can occupy more charge states, effectively diluting the signal for any given form of a peptide or

protein. Thus, sensitivity is sacrificed for greater connectivity of PTMs in these experiments. Table 1.1 summarizes the relative advantages and disadvantages of each approach.

Table 1.1. Comparison of mass spectrometry techniques.

	Scope	Advantages	Disadvantages

Bottom-up	Small peptides	- Best sensitivity
		- Easiest analysis
			- Lose connectivity of many PTMs
			- Paired with CID:
			  - Lose labile PTMs
			  - Non-random backbone cleavage

Middle-down	Medium peptides	- Better connectivity than bottom-up
		- Better sensitivity than top-down
		- Paired with ETD:
		  - Retain labile PTMs
		  - Even backbone cleavage
			- Lose connectivity of some PTMs
			- Worse sensitivity than bottom-up

Top-down	Intact proteins	- Complete connectivity of PTMs
		- Paired with ETD
		  - Retain labile PTMs
		  - Even backbone cleavage
			- Difficult data analysis
			- Worst sensitivity

Table adapted from (Karch et al., 2013).

1.2.3. Methods to study the structure and dynamic properties of histone proteins

Researchers have long been interested in understanding the structure and dynamics of histones due to their critical importance in the nucleus. The stability of the nucleosome is tied to many critical nuclear processes, such as maintaining higher order chromatin structure and transcription. For example, nucleosomes located in or near protomoter regions and other transcriptionally active chromatin regions have been found be less stable in many organisms including yeast (Moyle-Heyrman et al., 2013), plants (Vera et al., 2014), *Drosophilia* (Henikoff et al., 2009), and human cells (Ishii et al., 2015; Jin et al., 2009; Kubik et al., 2015). Therefore, understanding the stability and structure of the nucleosome as well as sub-nucleosomal histone

complexes (such as H3/H4 tetramers) in various biological contexts is critical to more clearly elucidate their role in regulating nuclear processes.

To date, there have been many studies aimed to understand the structure and dynamic properties of sub-nucleosomal histone complexes and nucleosomes using a variety of methods. In 1997, Luger et al. solved the crystal structure of the nucleosome to 2.8 angstrom resolution, representing the first global highly-resolved view of nucleosome structure (Luger et al., 1997). Since then, several nucleosome structures have been solved from various organisms. Notably, the crystal structure of the nucleosome with linker histone H1 bound was recently solved along with an 11 angstrom cryo-EM structure of an H1-containing chromatin fiber (Song et al., 2014; Zhou et al., 2015). However, the tail domains in each of these structures were unable to be determined due to low electron density in the corresponding regions. These structural studies are incredibly important to lend insight into the function of these macromolecules. However, one major limitation of crystallographic and cryo-EM studies is that they provide a static structure of the protein of interest. They therefore cannot be used to measure dynamic properties or shed insight into the stability of a given structure within the protein or complex.

Similarly, nuclear magnetic resonance spectroscopy (NMR) has been used to study histone structures. For example, Moriwaki et al. determined the structure of the H2A/H2B dimer using NMR (Moriwaki et al., 2016). Additionally, other studies have used NMR to map the binding interface of a nucleosome-interacting protein (e.g. Kato et al., 2011; Peña et al., 2009). Despite the great utility of these studies, NMR is limited by the size requirements of the analyte of interest. Nucleosomes are too large to be studied by NMR, and so separate experiments must be done to study each component of a nucleosome separately. Furthermore, the heavy labeling required for NMR experiments is very costly and may be prohibitive to study nucleosomes.

Molecular dynamics studies have also been used to probe histone structure and dynamics (e.g. Flanagan and Brown, 2016; Kenzaki and Takada, 2015; Korolev et al., 2014; Li and Kono, 2016). However, molecular dynamics simulations require a lot of computer power and

therefore are often very short simulations or do not contain the entire nucleosomal complex. These studies are also often very difficult to validate experimentally, and are very dependent on the exact parameters used in the study.

There have also been studies using FRET to monitor stability of specific structures within the nucleosome (Falk et al., 2016; Lehmann et al., 2017; Nakaoka et al., 2016; Sasaki and Yoshida, 2014). However, FRET studies can only measure the distance between an acceptor and donor fluorophore and can therefore be used to study one structure at a time and cannot provide a global analysis of histone dynamics.

Hydrogen deuterium exchange (HDX) studies have also been employed to study histone dynamics. In an HDX experiment, proteins of interest are solvated in heavy water (deuterium oxide, D_2O) over a course of time. Amide protons on the backbone of the protein can then exchange for deuterium atoms from the solvent. However, amide protons may be protected from exchange if they are not readily accessible by solvent or if they are participating in a hydrogen bond, as is the case for secondary structures. However, transient unfolding events can break hydrogen bonds and lead to deuteration of the protein backbone. The rate of deuterium incorporation therefore acts as a proxy for protein structure and stability, with faster rates indicating less stable structures (Englander, 2006). HDX can also be used to map binding interfaces with small molecule ligands, other proteins, or nucleic acids as amino acids involved in hydrogen binding at the interface will be more protected from exchange. Deuterium incorporation can be measured by mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR). A more thorough introduction to HDX is given in Chapter 4.1.

HDX has been used to study canonical $(H3/H4)_2$ heterotetramers (Black et al., 2004), H2A/H2B dimers (D'Arcy et al., 2013; Moriwaki et al., 2016), and nucleosomes (Black et al., 2007). It has also been employed to study histone variants (Bassett et al., 2012; Black et al., 2004, 2007; DeNizio et al., 2014; Falk et al., 2015; Guo et al., 2017; Panchenko et al., 2011), nucleosomal arrays (Panchenko et al., 2011), and histone interaction with binding partners, such

as chaperones (Bassett et al., 2012; D'Arcy et al., 2013; DeNizio et al., 2014; Falk et al., 2015; Guo et al., 2017).

The aforementioned techniques and studies have provided great insight into the structure and dynamic properties of histone complexes and nucleosomes. However, each of these techniques is unable to rigorously and globally address the structure and dynamics of the histone tail domains that protrude from the nucleosomal surface, likely because they are highly flexible regions of the protein. For example, there is not enough electron density for the tail domains in crystallography studies, and so the tail domains have to be modeled and mapped to the structure. Furthermore, as will be discussed in more detail in Chapter 4.1, traditional HDX techniques are not well-equipped to analyze tail domains due to the fact that pepsin digestion generates extremely long peptides in the tail domains that prevents resolved localization of deuterium content. As such, more work remains to be done to understand nucleosome structure and dynamics. This insight will shed light on basic nuclear processes and mechanisms.

CHAPTER 2: Development and evaluation of mass-spectrometry based methods to study histone PTMs

2.1: Introduction

Histone proteins are among the most post-translationally modified proteins in the cell, with a majority of the post-translational modifications (PTMs) occurring on the N-terminal tail domains of the histones, although some occur on the histone globular core as well. Histone H3 and H4 undergo the most extensive modification, and aberrant regulation of these PTMs has been linked to many diseases, including cancer (Portela and Esteller, 2010).

As such, it is of great interest in the chromatin field to be able to accurately and robustly identify and quantify histone PTMs in a reproducible manner. Mass spectrometry (MS) has emerged as a powerful tool to accomplish this task in an unbiased manner. MS techniques also have the advantage of being able to identify novel PTMs as well as the occurrence of several PTMs on a single peptide.

Most histone analyses using MS have employed data-dependent acquisition (DDA) methods (Figure 2.1B). In DDA methods, a full scan is obtained followed by a series of MS/MS scans of ions selected from the full scan. Usually, the ions are chosen sequentially by their abundance, with the most abundant ion chosen first (Aebersold and Mann, 2003). Then, label-free quantification is achieved at the full MS level by integrating the area under the LC peak of an extracted ion chromatogram corresponding to the monoisotopic peak of the ion distribution.

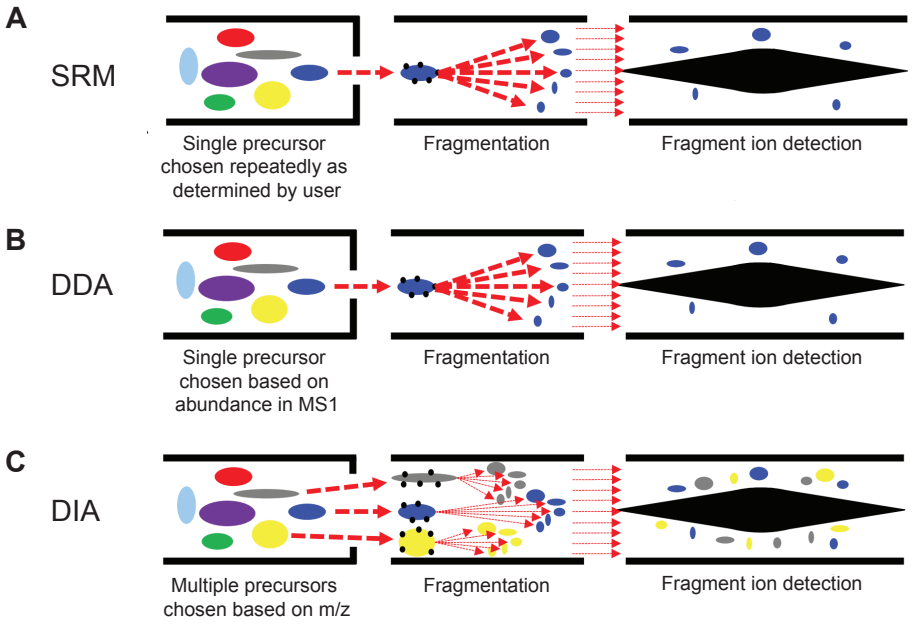


Figure 2.1. MS data acquisition methods. (A) Selected reaction monitoring (SRM) methods isolate a single precursor ion defined by the user for fragmentation across its entire elution range. (B) Data-dependent acquisition (DDA) methods isolate a single precursor ion for fragmentation based on its abundance in the MS1 spectrum. (C) Data-independent acquisition (DIA) methods selects all precursor ions in a given m/z window for fragmentation and measures the masses of the combined fragment ions in a single scan. Figure is modified from (Hu et al., 2016).

However, many histone peptides are isobaric, meaning that they have the same exact mass but different chemical composition. For example, the H3 9-17 peptide containing K9ac or K14ac have the same exact mass and co-elute, making it impossible to quantify these two peptides separately at the full MS level. To circumvent this issue, most histone MS studies incorporate a selected reaction monitoring (SRM) acquisition in the method whereby isobaric peptides of interest are targeted for fragmentation across their elution profile (Figure 2.1A) (Lin

and Garcia, 2012). SRM can also be used for quantification of non-isobaric peptides. Quantification is then performed on the MS/MS level using unique fragment ions (i.e. those that are different between the isobaric species). Accurate quantification cannot be performed based on a single MS/MS spectrum, but rather should be performed over the entirety of the peptide elution, and so each SRM analysis within the method must be tailored to the specific isobaric peptides of interest (Picotti and Aebersold, 2012). As such, SRM requires prior knowledge of retention times and therefore is heavily reliant on reproducible chromatography. Nonetheless, this DDA/SRM hybrid MS methodology has been very powerful for comprehensive PTM analysis.

Most histone PTM DDA/SRM MS studies to date have employed high resolution mass analyzers, mainly Orbitraps, rather than low-resolution mass analyzers, such as ion traps. The greater mass accuracy afforded by high-resolution instruments allows for the discrimination of nearly isobaric PTMs, acetylation (42.0106 Da) and tri-methylation (42.0470 Da), while low-resolution instruments are unable to resolve such a small mass difference (Yang et al., 2010).

However, other techniques can be used to expand the capabilities of low-resolution instruments to distinguish these PTMs. Utilizing MS/MS data in combination with retention time information, for example, has been used to distinguish acetylation and tri-methylation on low resolution instruments (Falick et al., 1993; Kim et al., 2002; Klammer et al., 2007; Yang et al., 2010; Zhang et al., 2004). For example, the Hunt lab performed reversed-phase (RP) HPLC-MS/MS on histone peptides. The peptides were separated over a 240 minute gradient, and it was found that the H3 peptide spanning residues 9-17 containing K14ac or K9me3 eluted 17 minutes apart, demonstrating the potential of using retention time differences to distinguish these modifications (Syka et al., 2004). The Freitas group subsequently utilized this difference in retention time to demonstrate that tri-methylaiton and acetylation can be easily distinguished on a low-resolution instrument. They validated their findings by running the same sample on a high-resolution instrument (Yang et al., 2010).

Other groups have used heavy labeling to differentiate nearly isobaric PTMs (Cao et al., 2013; Ren et al., 2007; Smith et al., 2003; Zee et al., 2010). Heavy isotope labeled methyl or acetyl donors can be used to increase the mass of the PTM, and consequently increase the mass difference between these PTMs, enabling them to be distinguished on low-resolution instruments. However, it is relatively time consuming to wait for the light modifications to turn over. Furthermore, it is rare to reach 100% incorporation of the heavy label, and use of heavy labeled reagents is relatively expensive. Additionally, this method is limited to cell culture and cannot be used on tissue or patient-derived samples.

Another technique that can be used to distinguish these modifications is spiking in heavy labeled synthetic peptides corresponding to the peptides that cannot be distinguished by low-resolution instruments. The heavy-labeled peptide will co-elute with its unlabeled counterpart, but will be distinguished by its unique mass imparted by the heavy labels. Lee et al. ustilized this method to demonstrate that a potentially novel modification (H2AT15ac) was falsely assigned (Lee et al., 2013).

During an MS/MS experiment, peptides can undergo fragmentation of the PTM itself in addition to the peptide backbone. If this fragmentation does not alter the charge of the ion, it is called a neutral loss. Tri-methylated peptides undergo signature neutral losses during collision induced dissociation (CID) fragmentation methods. The presence of these neutral losses can aid in assignment of a tri-methyl PTM over an acetyl mark (Afjehi-Sadat and Garcia, 2013; Zhang et al., 2004). Acetylated peptides do not experience any neutral losses; however, fragmentation of acetylated peptides does produce immonium ions, which can be detected at m/z 143.1 and used to verify the presence of an acetyl group on the precursor peptide (Kim et al., 2002; Zhang et al., 2004).

Given that low-resolution instruments are less expensive, easier to maintain, and are more ubiquitious than high-resolution instruments, we sought to determine if low-resolution mass spectrometers are capable of robust, accurate, and reproducible identification and quantification

of histone peptides and PTMs using DDA/SRM methodology (Karch et al., 2014). To achieve this goal, we compared the performance of two common mass spectrometer instruments in comprehensive histone PTM analysis: (1) a low-resolution linear ion trap, LTQ Velos Pro (Thermo) and (2) a high-resolution linear ion trap-Orbitrap hybrid instrument, LTQ-Orbitrap Velos Pro (Thermo). Most of the common modifications have a large enough mass difference to be confidently assigned on a low-resolution instrument, and so the main challenge in using low-resolution instruments for histone PTM analysis is distinguishing acetylation and trimethylation PTMs.

Given that the low-resolution LTQ Velos Pro can operate in four different scan modes, we first sought to determine which is optimal for histone PTM analysis based on the following criteria: (1) reproducibility of relative peptide abundance measurements, (2) resolution of MS1 scans for peptides in higher charge states, and (3) the number of MS1 and MS2 scans acquired per run. We found that each scan mode resulted in reproducible relative quantification measurements, but the Enhanced scan mode has the optimal trade-off between resolution and number of scans. We then compared the performance of the LTQ Velos Pro operating in Enhanced scan mode to the high-resolution Orbitrap Velos Pro using the same criteria listed above. We found that, as expected, the Orbitrap Velos Pro was able to distinguish peptides bearing acetylation and tri-methylation while the LTQ Velos Pro could not. However, we determined that the low-resolution LTQ Velos Pro was able to distinguish these modifications if retention time information was used or if heavy-labeled synthetic peptide standards were employed in the analysis. As such, we conclude that while high-resolution mass analyzers are more easily suited for histone PTM analysis, low-resolution mass analyzers are equally capable of accurate, robust, and reproducible identification and quantification of histone PTMs using DDA/SRM methodology (Karch et al., 2014).

While DDA/SRM hybrid methodology has been a very powerful tool to study histone PTM profiles, there are several drawbacks to this method. To perform this method, previous knowledge

about retention time and mass of co-eluting isobaric species is required. Furthermore, targeting masses for fragmentation reduces the duty cycle of the method, and fewer DDA MS/MS spectra can be obtained as a result, reducing the chances of identifying low-abundant species. Another drawback is that the data acquisition and consequently peptide identification is somewhat stochastic because precursor selection is dependent on elution time, presence of other abundant ions, and dynamic exclusion lists. This can result in poor reproducibility between runs as a different precursors can be selected for fragmentation. Furthermore, because target masses in SRM methods are defined by the user pre-run, data cannot be re-mined upon discovery of new PTMs or isobaric peptides (Bern et al., 2010; Sidoli et al., 2015a).

Data independent acquisition (DIA) methods has been shown to circumvent many of these issues (Gillet et al., 2012; Ma et al., 2009). In DIA, a full MS scan is acquired followed by a series of sequential MS/MS scans covering a relatively large m/z window (usually ~5-25 m/z) that step across a desired m/z range (Figure 2.1C, 2.2). Therefore, several species can be fragmented together at once, and the resulting MS/MS spectrum will contain fragment ions from multiple species (Figure 2.1C). Using smaller windows across the desired m/z range allows for greater sensitivity and also reduces the complexity of the resulting MS/MS spectra (Gillet et al., 2012). Label-free quantification of isobaric species is then performed on the MS/MS level using an SRM-like approach whereby an extracted ion chromatogram of a product ion is generated in silico using bioinformatics tools such as Peakview ®, Skyline (MacLean et al., 2010), OpenSWATH (Röst et al., 2014), or EpiProfile (Yuan et al., 2015). In order to perform this analysis, a spectral library of identified peptides is required. This library allows for the fragment ions from a MS/MS spectrum to be mapped to the appropriate precursor peak. Spectral libraries can be downloaded, manually programed, or built by the user from previous DDA experiments. The Yates group was the first to demonstrate the strength of DIA methodology by using this method to measure differences in protein expression in two stages of development in *Caenorhabditis elegans* (Venable et al., 2004).

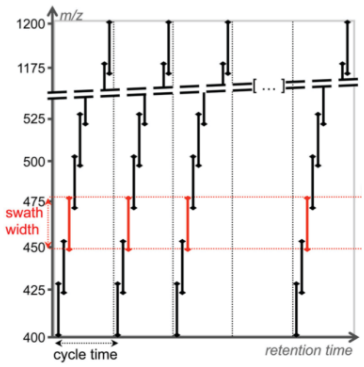


Figure 2.2. Principles of Data Independent Acquisition (DIA). In DIA methods, all ions in a given m/z window (here, 25 m/z as denoted by the double-headed arrows) are selected for fragmentation together. Sequential windows of ions are selected for fragmentation across a given m/z range (here, 400-1200 m/z for a total of 32 "steps"), allowing all ions to be selected for MS/MS. The cycle is repeated for the duration of the gradient, enabling all ions to be fragmented across their entire elution profiles. Figure is adapted from (Gillet et al., 2012).

Sciex was the first to automate DIA methodology by offering a DIA method called SWATH-MS (Sequential Window Acquisition of all THeoretical Mass Spectra) for their triple-TOF high-resolution mass spectrometer (Gillet et al., 2012; Hopfgartner et al., 2012). We have previously demonstrated the power of SWATH-MS to quantify and identify histone PTM profiles (Sidoli et al., 2015a). We compared the performance of DDA and SWATH for histone analysis using two model systems: (1) human embryonic stem cells (hESCs, strain H9), untreated (pluripotent) and treated with retinoic acid (RA) (differentiated), and (2) undifferentiated and differentiated mouse trophoblast stem cells (mTSCs). Three technical replicates were collected for each of the four samples with each method. The DDA data was used to demonstrate the high reproducibility of the chromatography and to identify histone peptides present in the sample. For

the SWATH-MS analysis, we focused primarily on histone H3 because it contains the most extensive and complicated PTM profiles. We demonstrated that SWATH-MS allowed for highly reproducible quantification of histone PTMs (average CV <8%) (Figure 2.3). The results also demonstrated that different fragment ions produce highly similar quantification results, indicating that accurate quantification can be achieved with any suite of fragment ions. Even extremely low-abundant species (<0.05%), such as H3K9me2S10phos and K9me3S10phos, were able to be quantified with high precision, illustrating the high sensitivity of SWATH-MS, with precise quantification being achieved over four orders of magnitude (Sidoli et al., 2015a).

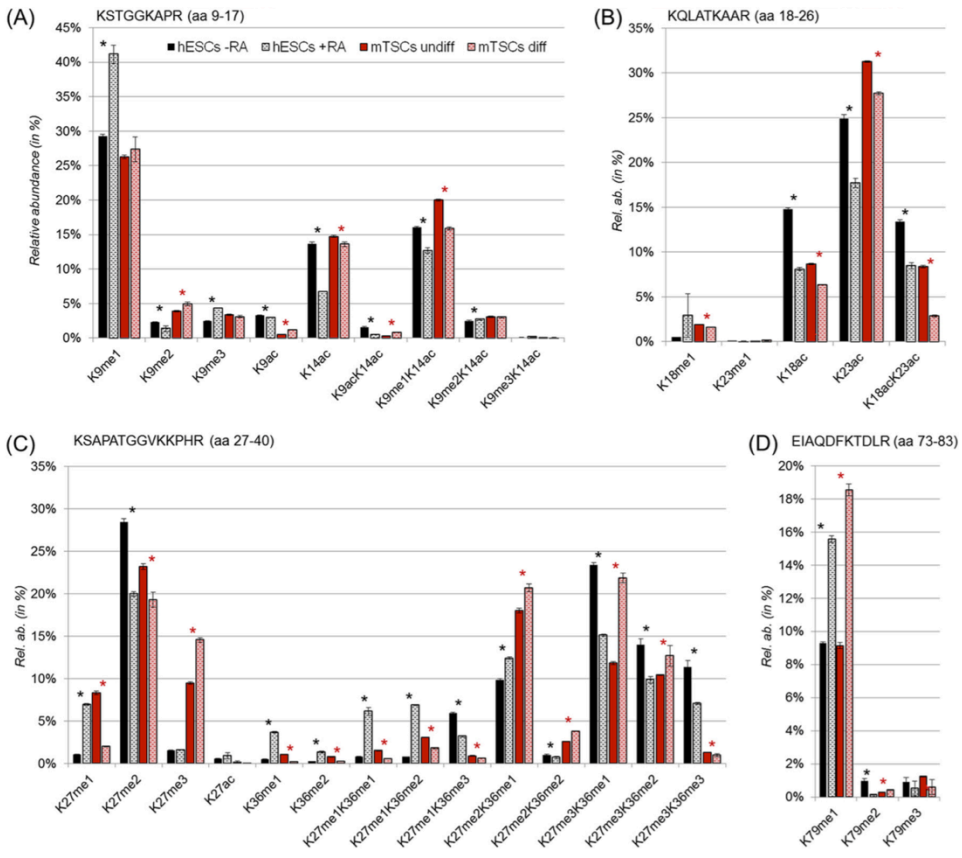


Figure 2.3. Histone H3 peptides quantified in hESCs and mTSCs during development. (A) Relative abundance of the peptide KSTGGKAPR (aa 9-17) in all of its methylated and acetylated

forms. hESCs are represented as black and gray bars, mTSCs as red and pink bars. The relative abundance in the four analyzed conditions is also displayed for the peptides (B) KQLATKAAR (aa 18-26), (C) KSAPATGGVKKPHR (aa 27-40), and (D) EIAQDFKTDLR (aa 73-83). The error bars represent standard deviation. The asterisks (*) represent a statistically significant ($p < 5\%$) difference between the two stages of development for hESCs (black asterisk) and mTSCs (red asterisk). Figure and caption taken from (Sidoli et al., 2015a).

The results of the SWATH-MS study and several other DIA studies from other groups demonstrate the power of DIA methodology for robust quantification of peptides in complex samples. However, given that histones are among the most extensively post-translationally modified proteins in cells and contain a very large number of isobaric co-eluting peptides, they are particularly challenging to study by MS. We therefore sought to optimize DIA methodology to maximize the number of peptides able to be analyzed in histone PTM analysis. Furthermore, we sought to determine if low-resolution MS and/or MS/MS are able to provide comparable results to high-resolution mass analyzers because low-resolution instruments tend to be more ubiquitous, cost-effective, and easier to maintain (Sidoli et al., 2015b). To this end, we extracted histones from mouse embryonic stem cells and prepared them by the traditional method of derivatization with propionic anhydride and digestion with trypsin. We analyzed this sample with a variety of DIA methods using both high- and low-resolution mass analyzers as well as the more traditional hybrid DDA/SRM analysis. We found that the quantification results were highly similar between the DIA methodology and the hybrid DDA/SRM methodology. We also demonstrate that the DIA method is more reliable because it is not as dependent on chromatography as the DDA method. We also show that DIA is more readily able to identify isobaric peptides containing different modification profiles and also allows for re-mining of data to quantify new isobaric co-eluting peptides. We also demonstrate that DIA does not require high-resolution, making it a more convenient and reliable approach compared to the traditional DDA/SRM methodology.

2.2: Results

2.2.1: High-resolution is not a strict requirement for characterization and quantification of histone PTMs using a DDA/SRM hybrid MS approach

2.2.1.1: Comparison of LTQ Scan Modes

The LTQ Velos Pro can operate in four different scan modes: Turbo, Normal, Enhanced, and Zoom. Each scan mode has a different scan rate (Table 2.1). Scan rate affects the reproducibility of peptide abundance measurements, resolution, and the number of scans per run. Each of these properties is crucial for MS analysis of histone PTMs, and we therefore sought to determine which mode is optimal for histone PTM analysis.

Table 2.1. Scan rate information for linear ion trap scan modes

Scan Mode	Scan rate (Da/s)	MS/MS per duty cycle	Av MS1 per run[a]	Av MS2 per run[a]	Rel # of scans per run
Zoom	**1,111**	**6**	**1,825.3**	**10,952**	**0.81**
Enhanced	**5,000**	**6**	**2,179.5**	**13,077**	**0.97**
Normal	**16,666**	**6**	**2,256.7**	**13,540**	**1.00**
Turbo	**125,000**	**6**	**2,373.3**	**14,240**	**1.05**

[a] Based on a 76 min gradient (2%B for 1 min; 2% to 30% B in 55 min; 30% to 98%B in 15 min; 98%B for 10 min; 98% to 2% B in 30 s; 2% B for 9.5 min) at 250 nL/min flow rate using an Eksigent NanoLC Ultra loading pump. Table taken from (Karch et al., 2014).

To accomplish this goal, we decided to analyze HPLC-purified histone H4 in triplicate in each scan mode. The histone was extracted from HeLa cells treated with butyrate, a histone deacetylase inhibitor. By blocking deacetylase activity, this treatment results in more pronounced and combinatorial acetylation patterns. The sample was derivatized with propionic anhydride and digested with trypsin prior to analysis by MS. The same sample was run in triplicate for each scan

mode using the same analytical column and HPLC gradients within 2 days to reduce variance due to instrument setup.

We first investigated the reproducibility of relative abundance measurements an example H4 peptide (amino acids 4-17, sequence: GKGGKGLGKGGAKR) in all four scan modes. Each of the four lysine residues in the peptide can be acetylated. Relative abundances are calculated by integrating the area under the curve of the extracted ion chromatogram (XIC) of the monoisotopic peak of the distribution and dividing it by the total abundance of that peptide in all of its modified forms and charge states. We found that each scan mode resulted in a relatively small standard deviation, indicating that each scan mode generates reproducible results (Figure 2.4). However, Turbo scan mode has significantly different relative abundance values than the other scan modes, indicating that this mode may be less accurate than the other three modes. This inaccuracy could be due to the presence of multiple species in a given m/z window that cannot be resolved in this mode. The tri- and tetra-acetylated peptides are generally lower in abundance and therefore have higher standard deviations, as is typical for low abundant species.

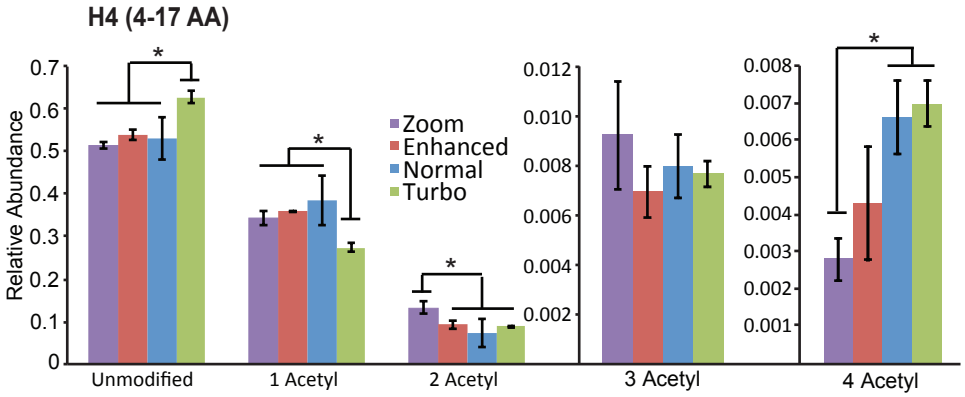


Figure 2.4. Reproducibility of relative peptide abundance measurements for H4(4-17AA) peptides on each LTQ Velos Pro scan mode. The relative abundance of each modified form of the peptide was calculated from three technical replicates of the same purified H4 sample. Error

bars represent standard deviation from average relative peptide abundances. *P < 0.05. Figure and caption taken from (Karch et al., 2014).

We next evaluated the resolution achieved for each scan mode, shown in Figure 2.5A for an example peptide, di-acetylated H4 4-17 +2 ($[M + 2H]^{2+}$ = 761.939). The results demonstrate that Turbo cannot resolve the isotopes of this example peptide and therefore cannot determine the charge state. The other scan modes, Normal, Enhanced, and Zoom, have sufficient resolution to determine the charge state of the peptide ion. Determining the charge state is critical for determining the parent mass of the ion for identification.

However, as charge state increases, it becomes more difficult to determine the charge at a given resolution because the isotopes are closer together in m/z. Propionylated tryptic histone peptides generally occupy lower charge states, +1 to +3. We examined the resolution obtained for the same example peptide in +3 charge state ($[M + 3H]^{3+}$ = 508.295) (Figure 2.5B). The results demonstrate that Zoom and Enhanced scan mode can resolve the peptide while Normal and Turbo scan modes cannot fully resolve the peptide. We therefore concluded that Normal and Turbo scan modes are not adequate for histone PTM analysis. Figure 2.5C displays the relationship between resolution and scan rate for this example peptide.

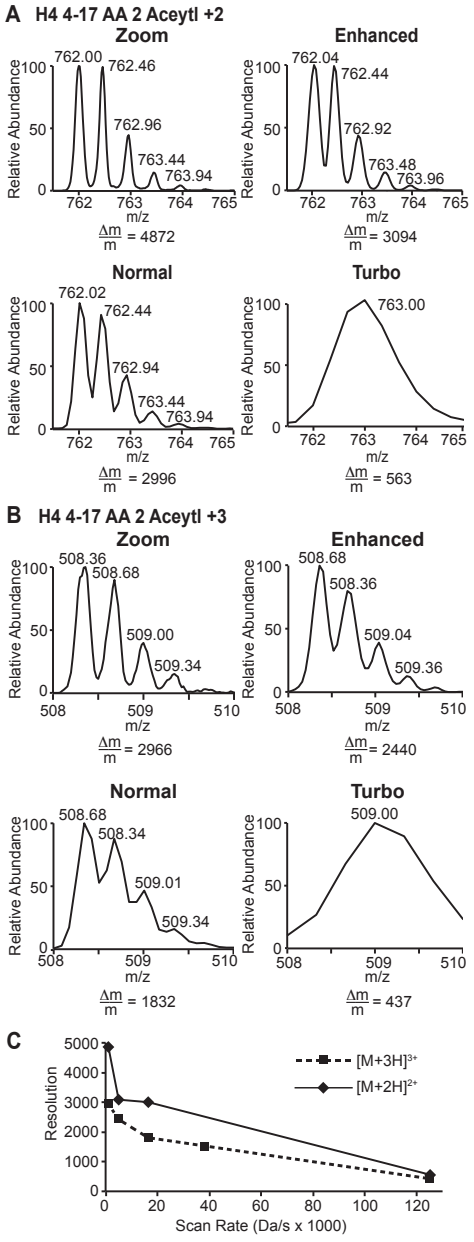


Figure 2.5. Resolution of di-acetylated H4 4-17 on each LTQ Velos Pro scan mode. Mass spectra of (A) doubly charged ($[M+2H]^{2+}$ = 761.939 m/z) and (B) triply charged ($[M+3H]^{3+}$ = 508.295 m/z) peptides are displayed. The resolution (Δm/m) of the monoisotopic peak is indicated. (C) Resolution of the peptides from (A) and (B) as a function of scan rate. Figure and caption taken from (Karch et al., 2014).

In addition to altering the resolution of the data, scan rate also affects the number of scans that can be collected in the analysis. Zoom, although it has the highest resolution, collects many fewer full MS and MS/MS spectra compared to the other scan modes (Table 2.1). Many histone modifications have low abundances, such as K18me1 and K23me1, so it may be beneficial for analysis to sacrifice higher resolution for an increased number of scans to increase the chances of selecting peptides with lower abundance to be selected for MS/MS analysis.

Based on these data, we recommend using Enhanced scan mode for histone PTM analysis. This mode offers reproducible quantification, sufficient resolution to resolve isotopes of peptides in +2 and +3 charge states, and about 20% more MS/MS scans compared to Zoom mode.

2.2.1.2: Comparison of LTQ and Orbitrap Performance

We next sought to compare the performance of a low-resolution LTQ Velos pro instrument with a high-resolution Orbitrap Velos Pro hybrid instrument to determine if low-resolution instruments can provide robust and accurate analysis of histone PTMs. We used a single sample of total acid-extracted histones from butyrate-treated HeLa cells. As a histone deacetylase inhibitor, butyrate increases the complexity of acetylation profiles, making it a more challenging sample to analyze. The sample was derivatized with propionic anhydride and digested with trypsin. Heavy-labeled synthetic peptides (also propionylated) were spiked into the histone sample to aid in identification of endogenous modified histone peptides. We analyzed three technical replicates of this sample on both instruments with the same analytical column, instrument method, and HPLC gradient to minimize variance due to instrumentation.

We compared the performance of the LTQ Velos Pro and Orbitrap Velos Pro based on scan rate, reproducibility of peptide abundance measurements, resolution, and mass accuracy.

The LTQ Velos Pro was operated in Enhanced scan mode as this was determined to be optimal for histone PTM analysis (Chapter 2.2.1.1).

One drawback of Orbitrap mass analyzers is that the scan rate is inherently slower than that of linear ion traps (Table 2.2). The Orbitrap Velos Pro collected an average of 73 full MS scans and 8,784 MS/MS scans, while the LTQ Velos Pro collected 1,189 and 11,703 scans, respectively. Therefore, the linear ion trap collects approximately 33% more scans than the Orbitrap. These results highlight the fact that while high resolution Orbitraps allow for greater mass accuracy and resolution, low-resolution linear ion traps collect a significantly larger amount of data even when operating in one of the slower scan mode options. This trade-off of reduced accuracy for increased numbers of scans could be desirable for identifying low-level histone PTMs.

Table 2.2: Scan rate information for Ion Trap and Orbitrap mass analyzers.

Detector	Average MS1 per run	Average MS2 per run	Relative # of scans per run
Ion Trap (Enhanced)	**1,189**	**11,703**	**1.33**
Orbitrap	**873**	**8,784**	**1.00**

*Note- Ion trap chromatography conditions are same as listed in Table 2.1. The Orbitrap data is based on a 66 minute gradient that omits the final equilibration step (2%B for 1 minute; 2% to 30% B in 55 minutes; 30 to 98%B in 15 minutes; 98% B for 10 minutes) at 250 nL/min flow rate using a Thermo Easy NanoLC HPLC. Table taken from (Karch et al., 2014).

As mentioned previously, it is important that the instrument used for analysis provides highly reproducible abundance measurements so that changes in abundance can be attributed to the sample rather than variations in instrument performance. Therefore, we sought to compare the reproducibility of abundance measurements between the LTQ Velos Pro and the Orbitrap Velos Pro using three technical replicates of the same histone sample on each instrument. H3 and H4 have the most complex PTM profiles and are the most difficult histones to analyze, so we chose to only quantify PTMs found on those histones. The results are displayed in Figure 2.6

and demonstrate that both instruments have a very small standard deviation, indicating that both instruments provide highly reproducible data. Furthermore, the standard deviations obtained on each instrument are very similar to each other ($P = 0.09$; paired t-test comparing standard deviations between the two instruments), indicating that the degree of reproducibility is very similar on both instruments.

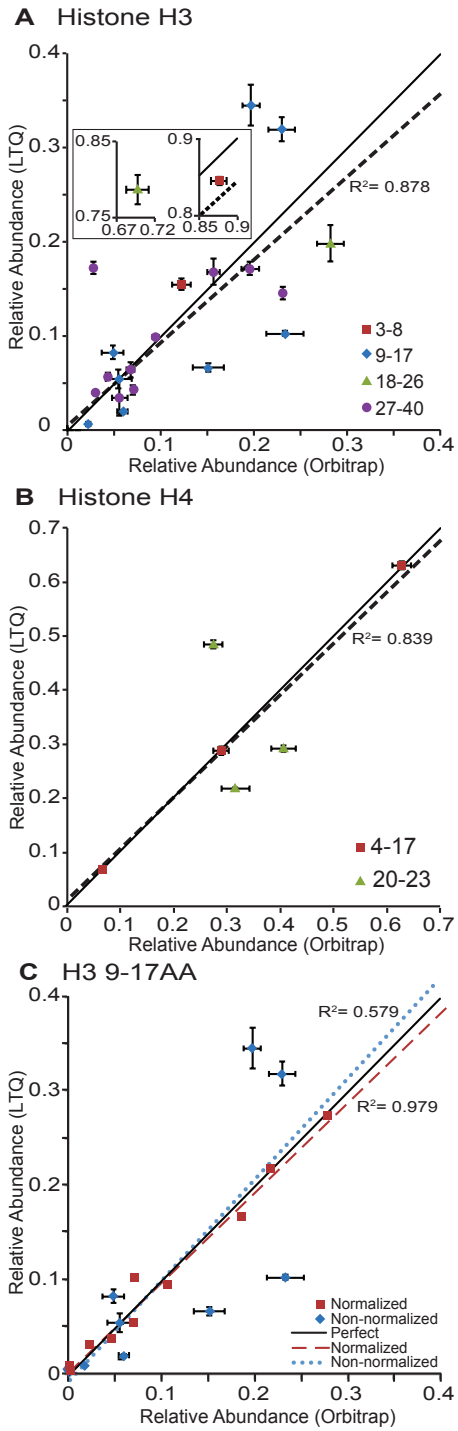


Figure 2.6. Reproducibility of relative peptide abundance measurements obtained on the LTQ Velos Pro and the Orbitrap Velos Pro. (A and B) Each point represents a particular modified form of either (A) an H3 peptide or (B) an H4 peptide, while the color of the point

indicates the identity of the peptide. The black line indicates perfect correlation between relative peptide abundance values on the two instruments. The dashed line is a linear regression for all data points on the plot (Pearson: (A) R^2 = 0.878 (B) R^2 = 0.839 Spearman: (A) R^2 = 0.853 (B) R^2 = 0.883). The modified peptides shown in the figure include: (A) 3-8: unmodified, K4me1; 9-17: unmodified, K9me1, K9me1K14ac, K9me3K14ac, K9me2, K9me3, K9 or K14ac, K9me2K14ac; 18-26: unmodified, K18 or K23 ac; 27-40: unmodified, K36me1, K36me2, K27me1, K27me2, K27me3, K27me2K36me1, K27me1K36me2, K27me1K36me3, K27me1K36me1 (B) 4-17: unmodified, mono-, and di-acetylated; 20-23: unmodified, K20me1, K20me2. (C) Normalization of endogenous relative peptide abundances to synthetic peptide standards. Each point represents a particular modified form of the H3 9-17 peptide, listed in (B). The blue points represent the data before normalization as shown in (B), and the red points represent the data after normalization to the synthetic peptide standards. The solid black line represents perfect correlation. The dashed blue line is a linear regression fit of the data before normalization (R^2 = 0.579), and the red dotted line is a linear regression fit of the data after normalization (R^2 = 0.979). Figure and caption taken from (Karch et al., 2014).

We next sought to determine if the instruments obtained similar relative abundance measurements. Since we used the same sample for each analysis, we would expect the results to be highly similar. The results indicate that in general this is true, but some calculated values differ between the two instruments (Figure 2.6A and 2.6B, as shown by deviation from the black line that indicates perfect correlation). A linear regression fit to all data points for H3 or H4 peptides shows a high degree of correlation between the relative abundance values obtained on each instrument (Pearson, R^2 = 0.878 for H3 and R^2 = 0.834 for H4; Spearman, R^2 = 0.853 for H3 and R^2 = 0.883 for H4). These results demonstrate that the instruments do produce highly similar abundance measurements. Also, given that the abundance measurements are relative, an error

in the abundance measurement of one peptide will affect the relative abundance of all other forms of that peptide.

The overall correlation of peptide abundance measurements between the LTQ Velos Pro and the Orbitrap Velos Pro are relatively high; however, some peptides do not correlate well between the two instruments. The peptide with the lowest correlation is the H3 9-17 peptide (Pearson: $R^2 = 0.579$) (Figure 2.6C). Low correlation is likely due to the major differences in how the instruments collect data. There were no common properties between modifications or peptides that did not correlate well, likely because one difference in peptide abundance measurements affects all other relative abundance calculations for that peptide.

We used the synthetic peptide library to correct for differences in instrument data acquisition. The synthetic peptide standards were added in equal concentrations. Therefore, the relative peptide abundance measurements of endogenous peptides can be normalized to those of the synthetic peptide standards because they are all present in exactly the same amount. To normalize the data, a normalization factor was calculated for each form of the peptide by dividing the expected relative abundance measurements of the synthetic peptides by the observed relative abundance measurement of the synthetic peptide. The raw abundance values of the endogenous peptides were averaged and multiplied by the normalization factor. The relative peptide abundances were calculated from these corrected raw abundance values. Our group has previously used this method to correct for differences in ionization efficiencies between differently modified histone peptides for quantification (Lin et al., 2014). After normalization, the relative abundance measurements obtained on the two instruments are very highly correlated (Pearson: $R^2 = 0.979$) (Figure 2.6C). This result demonstrates that the low-resolution LTQ Velos Pro can quantify histone PTM abundances as accurately as the high-resolution Orbitrap Velos Pro instrument.

We next sought to compare the mass accuracy of two instruments. To this end, we analyzed two example peptides, H3 18-26 and H4 4-17 in all of their modified forms to determine

what mass tolerance can be used to identify peptides. The results show that the mass accuracy for the LTQ Velos Pro ranges from 53.0 to 193 ppm, with an average of 129 ppm. On the other hand, the mass accuracy of the Orbitrap Velos Pro ranges from 0.00 to 6.62 ppm, with an average of 2.01 ppm (Table 2.3). Therefore, the Orbitrap Velos Pro can distinguish acetylation and tri-methylation using a mass tolerance of 10 ppm, while the LTQ Velos Pro, using a mass tolerance of 150 ppm, cannot distinguish these modifications. Figure 2.7 shows the extracted ion chromatograms of selected modified forms of the histone H3 9-17 peptide. On the Orbitrap Velos Pro, the tri-methylated and acetylated peptide (row 4 and 5, respectively) can be definitively assigned to the single peak, while the LTQ Velos Pro cannot resolve them, as seen by two major peaks present in both chromatograms. One of these peaks is the acetylated peptide while the other is tri-methylated. The resolution of the LTQ Velos Pro is high enough to distinguish all other common PTMs, such as mono- and di-methylation and peptides containing several PTMs such as K9me1K14ac (Figure 2.7, rows 1-3).

Table 2.3. Mass accuracy information of example peptides for Linear Ion Trap and Orbitrap.

		Linear Ion Trap			Orbitrap		
Peptide	Modifications	Observed m/z	Theoretical m/z	Error Δppm	Observed m/z	Theoretical m/z	Error Δppm
H3 18-26	None	577.93	577.85	138	577.849	577.849	0.00
	K23me1	584.90	584.86	120	584.857	584.857	0.00
	K18me1	584.93	584.86	120	584.857	584.857	0.00
	K18 or K23ac	570.93	570.84	158	570.842	570.841	1.75
	K18acK23ac	563.93	563.83	177	563.832	563.833	1.18
H4 4-17	None	776.10	775.95	193	775.955	775.955	0.00
	1 Ac	769.02	768.95	91.0	768.948	768.947	1.30
	2 Ac	762.02	761.94	105	761.934	761.939	6.56
	3 Ac	754.97	754.93	53.0	754.926	754.931	6.62
	4 Ac	748.02	747.92	134	747.924	747.922	2.67
			Average	129		Average	2.01

Table taken from (Karch et al., 2014).

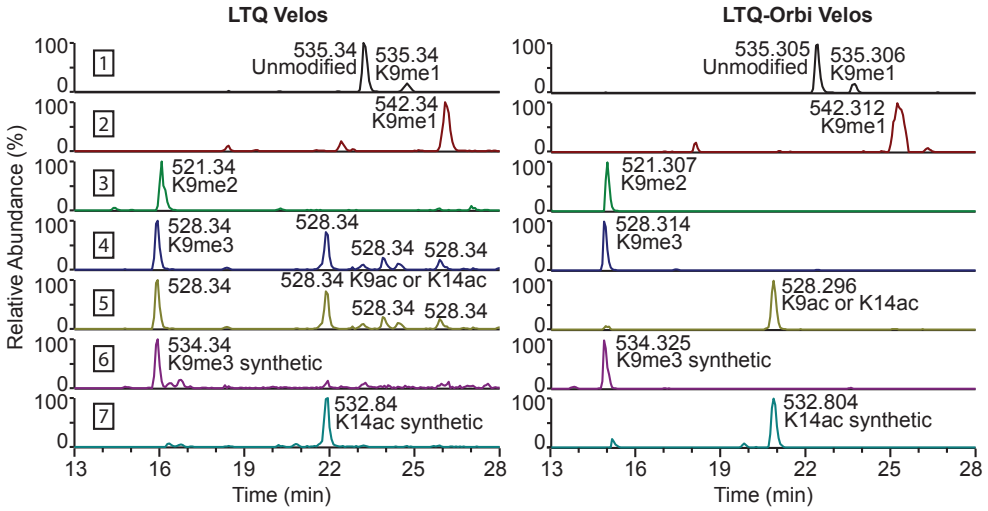


Figure 2.7. Chromatographic information for H3 9-17 obtained on LTQ Velos Pro or LTQ-Orbitrap Velos Pro. The mass tolerance used for selection is 150 ppm for the LTQ Velos Pro and 10 ppm for the LTQ-Orbitrap Velos Pro. In each row, labeled 1 to 7, the mass of the peptide bearing a specific modification was specified. Rows 1 to 5 represent endogenous peptides while rows 6 and 7 represent heavy-labeled synthetic peptides. Figure taken from (Karch et al., 2014).

The LTQ Velos Pro can still be used to distinguish acetylation and tri-methylation, however, using high-resolution MS/MS data and relative retention time information as shown by the Freitas group. However, this study did not use propionic anhydride derivatization and also did not perform quantification. Therefore, we decided to adapt this method for quantification of histone PTMs on both instruments. The retention time of acetylated peptides is later than methylated peptides because the acetyl groups are more hydrophobic than tri-methyl groups (Figure 2.7, row 4 versus 5). Derivatization of the peptides with propionic anhydride causes the unmodified peptide to be more hydrophobic than the di- and tri-methylated peptides. Di- and tri-methylation prevent derivatization with propionic anhydride and therefore do not contain that moiety at those modified residues. However, unmodified peptides will be propionylated at these residues, imparting a greater hydrophobicity than di- or tri-methyl groups. Therefore the

unmodified peptide is more hydrophobic than the di- and tri-methylated versions of the peptide, causing it to elute later (Figure 2.7, row 3 and 4 compared to 1). Although this relative retention time information can be used to distinguish peptides, validation of peak assignments with high-resolution instruments or synthetic peptides is ideal. Importantly, once the relative retention time is determined, validation will not be needed in subsequent experiments granted that the chromatographic set-up does not change.

The use of heavy-labeled synthetic peptides eliminates the need to perform high-resolution MS/MS for identification of peptides, and enables acetylation and tri-methylation to be distinguished based on retention times alone. Similar methods have been described by others, but there have not been studies using this method to distinguish nearly isobaric PTMs. Given that heavy isotopes, ^{13}C or ^{15}N, do not influence retention on C18 columns, the heavy-labeled synthetic peptides will have the same retention time as the unlabeled endogenous peptides. However, the heavy label will impart a unique mass to the synthetic peptide. Therefore, we can determine which ambiguous peak in the LTQ Velos Pro data contains a tri-methylated or acetylated peptide. This task can be accomplished by determining which ambiguous peak co-elutes with the synthetic peptide bearing the respective modification.

Figure 2.7 demonstrates how these synthetic peptides were used to differentiate tri-methylated and acetylated peptides. Rows 6 and 7 show the chromatograms for the H3 9-17 heavy-labeled synthetic peptides bearing K9me3 or K14ac, respectively. The K14ac synthetic peptide (row 7) co-elutes with the K14ac peptide in the endogenous trace (row 5), and the K9me3 synthetic peptide co-elutes with the corresponding endogenous K9me3 peptide peak (row 4). Therefore, elution times of the synthetic peptide can be used to assign ambiguous peaks in the endogenous peptides on the low-resolution LTQ Velos Pro.

2.2.2: Low-Resolution Data-Independent Acquisition Allows for Simplified and Fully Untargeted Analysis of Histone Modification

In this study, DIA was optimized for histone analysis and compared to traditional DDA/SRM hybrid methodology to determine if DIA is adequate for rigorous histone PTM analysis. Histones were extracted from mouse embryonic stem cells and prepared with the traditional propionic anhydride derivatization and trypinsization. This sample was used as a standard to compare sensitivity and precision of DDA and DIA methodologies in analysis of isobaric and non-isobaric histone peptides. The same sample was run in triplicate in each acquisition mode from the same vial within two days under the same nLC-MS setup to reduce variability from sample preparation and instrument setup. DDA runs were used as a reference for peptide quantification in the DIA runs. In DDA runs, specific co-eluting isobaric peptides were targeted for fragmentation across their elution profiles in an SRM experiment to enable quantification of those species, which include H3 9-17 peptide with 1 acetyl, H3 18-26 peptide with 1 acetyl, and H4 4-17 peptide with 1, 2, and 3 acetyl groups. This method has been used extensively in the past for quantification of histone PTM profiles.

2.2.2.1: Duty cycle of data-independent acquisition methods

We first sought to optimize the DIA methodology for peptide quantification. To this end, we tested 8 different DIA methods (Figure 2.8), all of which collect MS/MS scans at low resolution in the ion trap. We performed three technical replicates of the same sample for each method. We tried three different window sizes for MS/MS fragmentation, including 20, 40, and 50m/z, which results in 40, 20, and 16 MS/MS scans per duty cycle, respectively. Collecting such a large number of MS/MS scans increases the length of the duty cycle far beyond that of DDA methods. This reduces the number of full MS scans collected in DIA methods, resulting in a less defined chromatographic peak and consequently reduced accuracy of quantification on the full MS level.

To circumvent this issue, we designed four of the eight methods to include multiple full MS scans within a duty cycle to improve the definition of extracted ion chromatograms.

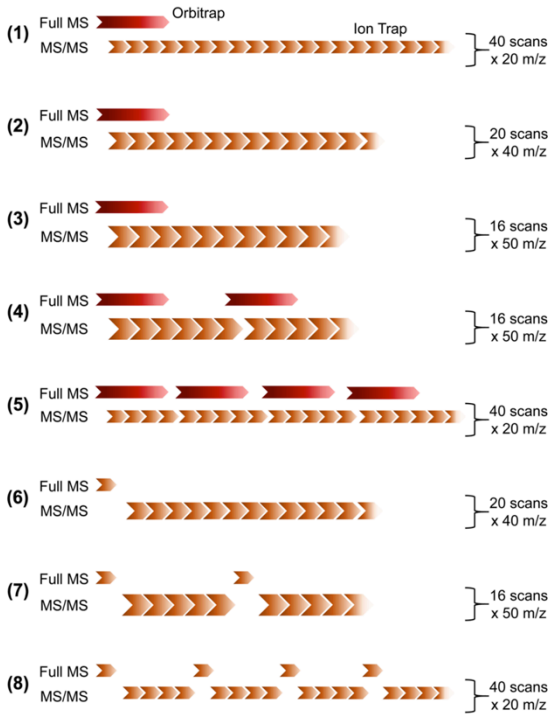


Figure 2.8. Data-independent acquisition (DIA) method experimental layouts. We evaluated eight different DIA methods in which we varied the MS/MS fragmentation window size (from 20 to 50 m/z), the number of full MS scan events per cycle (from 1 to 4), and the choice of the mass analyzer for the full MS scan (Orbitrap or the ion trap). The scan event in the Orbitrap mass analyzer is represented as red arrows, while the brown boxes represent the ion trap scans. The alignment between full scan and MS/MS events indicates whether the two scans are performed in parallel. The size of the boxes is roughly proportional to the scan time required for the scan. Figure and caption taken from (Sidoli et al., 2015b).

Given that low-resolution mass analyzers are more ubiquitious, cost-effective, and easier to maintain, we sought to test whether DIA could be performed in exclusively low-resolution mode while still providing comparable results to high-resolution analysis. Therefore, we included three methods in which the full MS scan is obtained in the low-resolution ion trap, using different MS/MS window sizes.

We determined the number of full MS and MS/MS scans for each method using RawMeat (Thermo), and found that DDA methods have a much faster duty cycle than the DIA methods, but that all of the DIA methods obtain a much larger number of MS/MS events, as expected (Figure 2.9). The results demonstrate that including more full MS scans in the duty cycle greatly increases the number of full MS scans obtained, while minimally affecting the number of MS/MS scans, indicating that the definition of MS/MS ion chromatograms would be mostly unchanged.

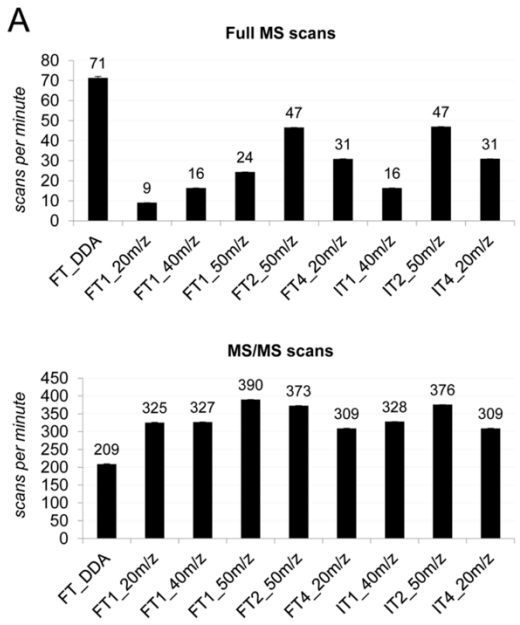


Figure 2.9. Scan frequency of the tested acquisition methods. Number of full MS scans (top) and MS/MS scans (bottom) per minute of the nine tested acquisition methods. The error bars represent standard deviation between three technical replicates. Figure and caption taken from (Sidoli et al., 2015b).

The high-resolution Orbitrap has a much longer scan time than the low-resolution ion trap, which theoretically could lead to longer duty cycles. In our case, we used a resolution of 120,000 at 200 m/z in the Orbitrap, which takes approximately 0.54 seconds, while a comparable scan in the ion trap (at much lower resolution) takes 0.09 seconds. However, scans in the Orbitrap can be parallelized with MS/MS scans in the ion trap (i.e. they can occur simultaneously because they occur in different mass analyzers). Thus, the only factor affecting duty cycle for full scans in the Orbitrap is the length of the injection time. We found that the frequency of MS and MS/MS scnas was highly comparable between methods using the ion trap and Orbitrap, indicating that these methods have a similar duty cycle speed. Together, we found that the methods FT2_50m/z and IT2_50m/z had the best definition of the full MS and MS/MS

chromatograms and the largest number of collected scans and are therefore best suited for histone PTM analysis.

2.2.2.2: Characterization of isobaric peptides and quantification of histone peptides

We evaluated the performance of the acquisition methods by testing identification and quantification of 111 peptides. Of these, 44 had isobaric forms that we discriminated and quantified at the MS/MS level, using the relative intensity of their unique fragment ions. This method allows for discrimination and separate quantification of differently modified peptides, for example H3 18-26 (sequence: KQLATKAAR), which can have an acetyl on K18 or K23.

Furthermore, DIA methods, in contrast to the targeted DDA method, allows for data re-mining after the LC-MS run. For example, histone H2A variants contain many isobaric, co-eluting peptides that are not targeted in traditional DDA methods. We were able to, in our DIA data, characterize 16 more isobaric peptides from H2A variants, including peptide 4-11 of canonical H2A, H2A.J, and H2A.X, as well as peptide 12-17 of canonical H2A in differently modified forms. Together, these results show that DIA is an effective acquisition method to confidently identify and discriminate isobaric peptides without the need to target them for fragmentation during the run.

We also sought to compare the DIA methods with the standard DDA methodology that has been used in the field for many years. We compared the relative abundances of all 111 quantified peptides between all DDA and DIA runs (including all 8 method variations), and observed a high degree of similarity between all the performed runs. To determine if the methods obtained similar peptide quantifications, we averaged the three replicates for each method and determined the degree of correlation between each analysis (Figure 2.10). The results demonstrate that each comparison yields a near perfect linear correlation and slope, demonstrating that each method yielded nearly identical results. The method with the most

variability was the IT4_20m/z, likely because many fewer MS/MS scans are collected in this method compared to others (Figure 2.10), leading to less well-defined MS/MS chromatograms and consequently less reliable quantification.

R^2 correlation

	FT_DDA	FT1_20m/z	FT1_40m/z	FT1_50m/z	FT2_50m/z	FT4_20m/z	IT1_40m/z	IT2_50m/z	IT4_20m/z
FT_DDA		0.98	0.99	0.99	0.98	0.98	0.95	0.97	0.93
FT1_20m/z	1.01		0.98	0.98	0.98	1.00	0.97	0.96	0.97
FT1_40m/z	0.97	0.95		1.00	1.00	0.98	0.95	0.97	0.92
FT1_50m/z	0.98	0.96	1.01		1.00	0.99	0.95	0.97	0.93
FT2_50m/z	0.97	0.95	1.01	1.00		0.98	0.95	0.97	0.92
FT4_20m/z	1.00	0.99	1.03	1.02	1.02		0.96	0.96	0.96
IT1_40m/z	1.00	0.99	1.02	1.01	1.02	1.00		0.93	0.97
IT2_50m/z	0.99	0.97	1.02	1.01	1.02	0.99	0.95		0.91
IT4_20m/z	0.99	0.99	1.01	1.00	1.01	1.00	0.99	0.97	

slope

Figure 2.10. Relative quantification of histone peptides and correlation between experimental methods. R^2 correlation (top, orange) and slope (bottom, green) from the linear regression between all experiments, after averaging the quantifications from the technical replicates. Figure and caption taken from (Sidoli et al., 2015b).

As discussed previously, the DIA method FT2_50m/z allowed for the highest number of full MS and MS/MS scans to be obtained compared to the other methods. This method also achieved very high correlation (R^2 = 0.98) and slope (0.97) with the standard DDA method and also has the lowest coefficient of variation between replicates, indicating high reproducibility and accuracy. The IT2_50m/z method performed similarly, with a correlation of 0.97 with DDA and FT2_50/mz methods and a nearly perfect slope with the two methods. These results demonstrate

that DIA methodology in low- and high-resolution instruments yields highly comparable results with the standard DDA method but while also allowing for quantification of more isobaric species.

2.3: Discussion

In this work, we sought to determine if low-resolution mass spectrometers are capable of robust and comprehensive PTM analysis using the DDA/SRM hybrid method, given that low-resolution instruments are less expensive, easier to maintain, and somewhat more ubiquitous compared to high-resolution instruments (Karch et al., 2014). To this end, we compared the performance of a low-resolution LTQ Velos Pro mass spectrometer with a high-resolution Orbitrap Velos Pro instrument in comprehensive histone PTM analysis. The results of these studies show, for the first time, that low-resolution mass analyzers, such as the LTQ Velos Pro, are indeed adequate for comprehensive PTM analysis by the DDA/SRM hybrid method, although data analysis is more easily facilitated on high-resolution instruments.

The low-resolution LTQ Velos Pro can operate in four scan modes with varying speeds: Turbo, Normal, Enhanced, and Zoom. We determined that Enhanced scan mode is optimal for histone PTM analysis based on reproducibility of abundance measurements, the number of scans acquired per run, and the resolution of +2 and +3 charged peptides, which are the charge states primarily occupied by propionylated tryptic histone peptides.

Since small changes in PTM abundances can have large biological implications, it is critical to run samples on an instrument with high reproducibility so that any observed differences can be attributes to biological phenomena rather than technical variance. We found that all scan modes provided highly reproducible data and are therefore suitable for PTM analysis based on the reproducibility criteria alone.

However, there were major differences between the scan modes in terms of resolution and scan rate. Obtaining adequate resolution is crucial for identification and quantification of histone PTMs. If the resolution is too low, the charge state cannot be determined and the peptide cannot be identified. The resolution of a linear ion trap is largely determined by the scan rate. Ion traps scan and record the m/z range by increasing the radio frequency (rf) voltage applied to the electrode over time. This allows ions of increasing m/z to be ejected, detected, and recorded sequentially. Fast scan rates have lower resolution because the ions of a given m/z do not have enough time to fully eject before the rf increases, causing peak widening (Wong and Cooks, 1997). This was observed in the results of this study as well. The fastest scan rate, Turbo, was not able to resolve the isotopes of +2 and +3 charged peptides, while the second fastest scan rate, Normal, was not able to completely resolve +3 charged peptides. The slower scan rates, Zoom and Enhanced, could resolve the +2 and +3 charge states and are therefore amenable to histone PTM analysis. However, Zoom, as the slowest scan rate, collects much fewer MS1 and MS2 scans, which could preclude analysis of some low-abundant species. Given that many histone PTMs have low abundances, Enhanced scan mode was chosen as the optimal mode for histone PTM analysis due to its ability to resolve isotopes while allowing for more scans to occur.

After determining the optimal scan mode for histone PTM analysis, we sought to compare the performances of the low-resolution LTQ Velos Pro and the high-resolution Orbitrap Velos Pro instruments to determine if low-resolution is adequate for histone PTM analysis using the DDA/SRM hybrid method. One of the main reasons high-resolution instruments have been at the forefront of histone PTM analysis is that they can resolve the mass difference between tri-methylation and acetylation modifications (42.0470 and 42.0106 Da, respectively). These PTMs are highly abundant on histone proteins, and so it is critical for analysis that they can be distinguished from each other.

Previous groups have distinguished these modifications on proteins using a combination of retention time information and MS/MS information (Yang et al., 2010). Furthermore, Krey et al.

demonstrated that low-resolution LTQ and LTQ Velos Pro instruments can quantify protein abundances as accurately as high-resolution Orbitrap instruments (Krey et al., 2013). These studies demonstrate the potential of low-resolution instruments for comprehensive analysis of histone PTM profiles. To our knowledge, there has not been a side-by-side comparison between low- and high-resolution instruments in their ability to perform PTM identification and quantification. We therefore chose to do this comparison on histone proteins, which have among the most complicated and varied PTM profiles of any Eukaryotic protein.

We compared the performance of the LTQ Velos Pro and Orbitrap Velos Pro based on number of scans per run, reproducibility of abundance measurements, and mass accuracy. Our comparison showed that both instruments yield highly reproducible abundance measurements, and that the calculated relative abundance measurements are highly similar (R^2 = 0.979) after correcting for differences in instrument acquisition through the use of synthetic peptide libraries. The correction factors only need be calculated one time, and can be applied to all analyses. The results also show that the low-resolution LTQ Velos Pro has a much lower mass accuracy than the Orbitrap Velos Pro (average of 129 and 2.01 ppm, respectively). Therefore, the LTQ Velos Pro cannot distinguish between acetylation and tri-methylation PTMs based on mass alone. The use of heavy-labeled synthetic petpides, however, can be used to distinguish these ambiguous peaks because they co-elute with the endogenous modification but have a unique mass. All together, these results show that although the LTQ Velos Pro is not able to resolve the small mass difference between acetylation and tri-methylation, other orthogonal lines of evidence can be used to accurately identify ambiguous peaks. Therefore, high-resolution is not a strict requirement for histone PTM analysis, but can facilitate analysis due its high mass accuracy.

However, there are several drawbacks to the traditional DDA/SRM hybrid methodology described here, including variability associated with the somewhat stochastic nature of precursor selection for MS/MS, the inability to discriminate isobaric species without targeting, and inability to re-mine data post-run. To overcome these issues, data-independent acquisition (DIA) strategies

have been developed. Targeting windows of m/z ranges for MS/MS ensures that all ions are selected for fragmentation across their elution profiles, enabling quantification of more co-eluting isobaric species compared to DDA methods.

Here, we demonstrate that DIA methods can robustly identify and quantify complex histone PTM profiles. Furthermore, we show that low-resolution instruments, which are more cost-effective and ubiquitous, perform as well as high-resolution instruments in DIA analysis. It is important to note that we performed this work on an Orbitrap Elite mass spectrometer (Thermo Scientific), and that other types of mass spectrometers may perform optimally with different MS/MS window sizes and number of full MS scans per duty cycle.

We found that performing more than one full MS scan per duty cycle increases the definition of the full MS extracted ion chromatogram and consequently improves quantification. The addition of more full MS scans to the duty cycle minimally affects the number of MS/MS scans (and consequently definition of MS/MS extracted chromatograms) for both low- and high-resolution acquisition methods. We also found that DDA and DIA methods perform highly accurate and reproducible quantification of co-eluting isobaric species; however, DIA methods allow for quantification of a greater number of these species because targeting is not required. Furthermore, we found that all DIA methods tested here, as well as the DDA/SRM hybrid method, performed highly similarly and yielded very similar results. Together, these results demonstrate that DIA methods are very powerful tools to study complex samples such as histones, providing robust and reproducible analysis of peptides, including many isobaric species. These methods could also be applied to analysis of other highly modified proteins.

2.4: Methods

2.4.1: Low-resolution DDA/SRM hybrid study

2.4.1.1: Cell culture and sample preparation

Hela S3 cells were grown in suspension as previously described (Thomas et al., 2006). Cells were harvested by pelleting cells at 1,000 rpm for 5 minutes. Media was removed and cells were washed with 20 mL sterile PBS. Nuclei were extracted as previously described using detergent (Lin and Garcia, 2012). Histones were purified using a salt extraction followed by an acid extraction. Isolated nuclei were resuspended in low salt buffer 10:1 buffer:pellet by volume (0.4M NaCl, 1mM DTT, 0.3mM AEBSF, and 10mM sodium butyrate) and incubated at 4˚C with shaking for 30 minutes. The nuclei were pelleted at 3,000xg for 5 minutes at 4˚C, and the supernatant was decanted. The pelleted was resuspended 5:1 buffer:pellet by volume in high salt buffer (2.5M NaCl, 1mM DTT, 0.3mM AEBSF, and 10mM sodium butyrate). An equal volume of cold 0.4N H_2SO_4 was added slowly and the nuclei were incubated at 4˚C with shaking for 2 hours. The nuclei were pelleted at 3,400xg for 5 minutes and the supernatant, which contains histone proteins, was moved to a new tube. Proteins were precipitated using TCA as previously described (Lin and Garcia, 2012). When specified, extracted histones were separated offline using reversed-phase high performance liquid chromatography (RP-HPLC) as previously described (Lin and Garcia, 2012).

To prepare samples for MS analysis, acid-extracted total histones or RP-HPLC-purified histone proteins were chemically derivatized with propionic anhydride and digested with trypsin as previously described (Lin and Garcia, 2012). Samples were desalted using homemade C18 stage-tip columns as previously described (Lin and Garcia, 2012). Samples were resuspended in 0.1% formic acid to 1 ug/uL for MS analysis.

The synthetic peptide library was created as described by Lin et al. (Lin et al., 2014). Briefly, 93 synthetic peptides corresponding to heavy-labeled versions of histone peptides were

synthesized. The most common PTM profiles were included in the library. Peptides were purified by RP-HPLC and resuspended in water to a final concentration of 0.27 pmol/uL/peptide. The peptides were propionylated and desalted as described previously (Lin and Garcia, 2012). The synthetic library was spiked into histone samples at a ratio of 100fmol synthetic peptides: 1ug histone.

2.4.1.2: Liquid chromatography mass spectrometry: LTQ Scan Mode Comparison Studies

A 75um i.d. fused silica microcapillary column was fritted and packed with Reprosil-pur C18 resin (3um, Dr. Maisch GmbH) in-house using a pressure bomb. A commercial fused silica emitter with a 10um tip (New Objective) was fitted to the column. Histone H4 samples (1.5ug) were loaded onto the column using an Eksigent NanoLC AS-2 autosampler and separated using an Eksigent NanoLC 2D Plus system HPLC delivering a 76 minute gradient (buffer A: 0.1% formic acid in water; buffer B: 0.1% formic acid in acetonitrile): 2% buffer B for 1 minute, 2-30% B in 55 minutes, 30-98% B in 15 minutes, 98% B for 10 minutes, 98-2% B 30 seconds, 2%B for 9.5 minutes at a flow rate of 250nL/min. The sample was eluted into a quadrupole ion trap (LTQ Velos Pro, Thermo Scientific) mass spectrometer operating in Zoom, Enhanced, Normal, or Turbo scan modes as specified. The instrument collected a full MS scan followed by 6 data-depended MS/MS scans of the 6 most abundant ions from the MS1 scan. CID (collision energy: 40, activation Q: 0.25, activation time: 10ms) was used to fragment ions. Three technical replicates were performed for each scan mode type.

2.4.1.3: Liquid chromatography mass spectrometry: LTQ Velos Pro/LTQ-Orbitrap Velos Pro Comparison studies

The same column and HPLC set-up was used for these studies as described for the LTQ scan mode comparison project (as described in Chapter 2.4.1.2). In this case, a single sample of acid-extracted total histone with synthetic peptide library spike-in was run on a linear ion trap (LTQ Velos Pro) and a linear ion trap-Orbitrap hybrid instrument (Orbitrap Velos Pro). Data acquisition was performed similarly for both instruments. Acquisition was separated into three segments, 14, 26, and 16 minutes long, respectively. In the first segment, a full MS scan was acquired followed by 9 data-dependent MS/MS scans based on abundance. In the second segment, a full MS scan was obtained followed by 5 targeted MS/MS scans: 528.30, 570.84, 754.93, 761.94, and 768.95 m/z. The targeted scans were followed by 5 MS/MS scans of the top 5 most abundant ions from the full MS scan. In the third segment, a full MS scan was acquired followed by 10 data-dependent MS/MS scans of the most abundant ions. On the Orbitrap Velos Pro instrument, all full MS scans were obtained in the Orbitrap (profile mode, resolution: 60,000 at m/z 400) and MS/MS scans were obtained in the ion trap. The ion trap was operated in Enhanced mode for both instruments. CID fragmentation (collision energy: 40; activation Q: 0.25, activation time: 10ms) was used in all cases. Three technical replicates were performed for each instrument.

2.4.1.4: Data analysis

We used EpiProfile software to identify and quantify histone PTMs for the LTQ-Orbitrap Velos Pro data (Yuan et al., 2015). The algorithm uses MS, MS/MS, and retention time information to identify peptides. It also provides relative abundance calculations for each identified peptide. Abundances are calculated by measuring the area under the curve of an extracted ion chromatogram corresponding to the monoisotopic peak of an ion of interest.

Relative abundance measurements are calculated by dividing the abundance of a particular modified peptide in all of its occupied charge states and dividing it by the total abundance of that peptide in all of its modified and unmodified forms. All data from the LTQ Velos Pro instrument were manually quantified using XCalibur and Excel software because they were not compatible with EpiProfile at the time of analysis (EpiProfile is now capable of analyzing low-resolution data). XCalibur Qual Broswer was used to view and calculate raw abundances of selected peptides. All occupied charge states were evaluated.

2.4.2: Low-resolution DIA study

2.4.2.1: Histone extraction and digestion

Mouse embryonic stem cells were grown using standard media and harvested as previously described (Thomas et al., 2006). Nuclei were isolated and histones were extracted using a standard acid extraction protocol as previously described, with some adjustments (Lin and Garcia, 2012). Briefly, nuclei were isolated with detergent, and histones were extracted with 0.2M H_2SO_4 for 2 hours and precipitated with 33% tricholoroacetic acid (TCA) overnight. Histone proteins were derivatized with propionic anhydride and trypsinized as previously described (Sidoli et al., 2016). Samples were desalted with home-made C18 stage-tip columns as previously described (Lin and Garcia, 2012; Sidoli et al., 2016).

2.4.2.2: Liquid chromatography mass spectrometry

The NanoLC was fitted with a 75um i.d. x 17cm Reprosil-Pur C18_AQ (3um; Dr. Maisch GmbH, Germany) nanocolumn using an EASY-nLC nano HPLC (Thermo Scientific, Odense, Denmark), packed in-house. Samples were analyzed using the following gradient (buffer A: 0.1% formic acid; buffer B: 95% acetonitrile, 0.1% formic acid): 2-28% B in 45 minutes, 28-80% B in 5

minutes, 80% B for 10 minutes at a constant flow rate of 300 nL/min. The nLC was coupled to an LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific, Bremen, Germany). In all DIA methods, a full scan was acquired in the Orbitrap (range: 300-1100 m/z, resolution 120,000 at 200 m/z, AGC target: 2 x 105) or in the ion trap (AGC target: 3 x 104, injection time limit: 30 or 60 msec). All MS/MS data was obtained using CID fragmentation (normalized collision energy: 35) and collected in centroid mode. For DDA experiments, acquisitions were obtained in three segments: (1) 14 min: MS/MS of the top 7 most abundant ions, (2) 27 min: targeted CID fragmentation of common isobaric species (H3 9-17 with 1 acetyl, H3 18-26 with 1 acetyl, and H4 4-17 with 1, 2, or 3 acetyl groups) followed by CID fragmentation of the top five most abundant ions, (3) 19 min: CID fragmentation of the top 10 most abundant ions. Raw files can be obtained on the Chorus database (https://chorusproject.org), project number 923.

2.4.2.3: Data analysis

A spectral library was manually generated using Skyline (MacLean et al., 2010), considering all of the commonly detected peptides in histone analysis as described previously (Yuan et al., 2015). Within Skyline, peak extraction was optimized manually using previous knowledge of peptide retention times. Relative quantification was achieved by summing the intensities of the modified peptide in all of its charge states and dividing it by the total intensity of that peptide in all of its modified forms and charge states. Intensities were obtained by integrating the area under the curve of the extracted ion chromatogram. For isobaric peptides, the relative ratio of two isobaric forms was estimated by averaging the ratio for each unique fragment ion with different masses between the species. Statistical reproducibility was determined by estimating the coefficient of variation and linear regression. EpiProfile was used to calculate the relative ratio of isobaric forms of the H4 peptide 4-17 with one acetyl group and 3 acetyl groups because EpiProfile is able to estimate relative ratios of isobaric peptides containing more than 2 different forms (Yuan et al.,

2015). The calculated relative ratio was used to determine how much of the precursor area is attributed to each isobaric form.

CHAPTER 3: Identification and Quantification of Histone ADP-Ribosylation Sites in Response to DNA damage.

3.1: Introduction

ADP-ribosylation is a post-translational modification (PTM) that occurs on a wide variety of proteins throughout the cell. ADP-ribosylation can occur as a single ADP-ribose unit (mono-ADP-ribosylation; MARylation) or polymers of ADP-ribose units (poly-ADP-ribosylation; PARylation). ADP-ribosylation is catalyzed by a group of enzymes called ADP-ribose transferases (ARTs), which use NAD+ as a cofactor to add ADP-ribose units to an acceptor side chain (Hassa et al., 2006) (Figure 3.1). It has been previously reported that ADP-ribosylation can occur on the side chains of Lys, Arg, Glu, Asp, and Ser residues (Altmeyer et al., 2009; Laing et al., 2011; Leidecker et al., 2016; McDonald and Moss, 1994; Rosenthal et al., 2015; Zhang et al., 2013).

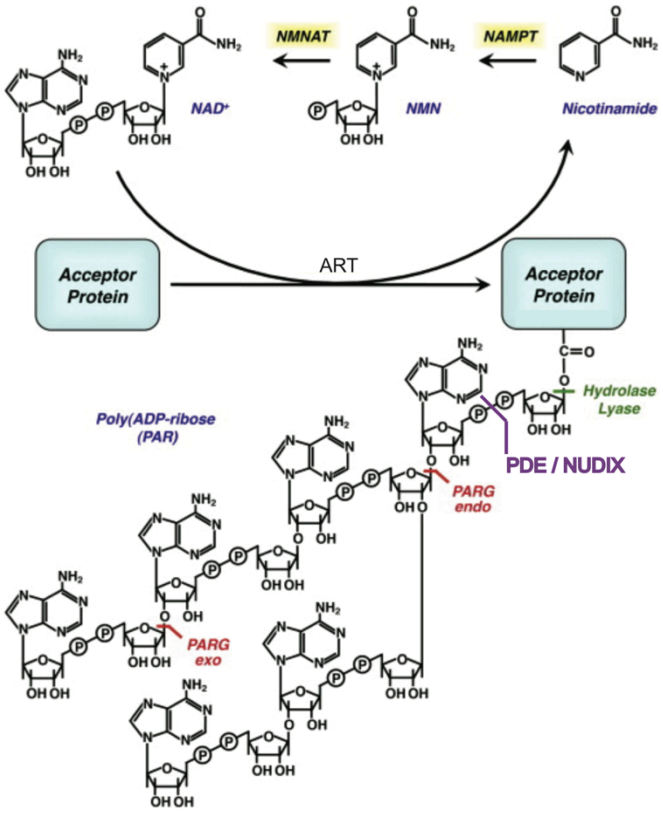


Figure 3.1. Biosynthesis of poly-ADP-ribosylation. The chemical structures of important metabolites and the PAR post-translational modification are shown. Enzymes responsible for degradation of PAR are also shown. Figure modified from (Krishnakumar and Kraus, 2010).

Addition of an ADP-ribose unit is a rather dramatic modification, with each monomer imparting a 541.0611 Da mass shift as well as the addition of two negative charges to the acceptor residue. Furthermore, ADP-ribosylation is a highly heterogeneous modification, with each acceptor site containing a varying number of ADP-ribose units that can exist in a linear chain or adopt a highly branched structure. Therefore, ADP-ribosylation, especially in the case of PARylation, can have a dramatic impact on protein chemistry and structure to alter protein function. Additionally, MARylation and PARylation can alter the interaction network of an acceptor

protein (Gibson and Kraus, 2012). Many “reader” domains have been identified that can bind to MARylated and PARylated proteins to mediate many cellular processes including differentiation, transcription, and stress response (Verheugd et al., 2016).

Of particular importance, nuclear ADP-ribosylation has been shown to play a vital role in DNA damage repair and maintenance of genome integrity. The most well-studied ARTs, PARP-1 and PARP-2, can bind directly to single and double stranded DNA lesions as the first step in the DNA repair process (Eustermann et al., 2015; Langelier et al., 2012). Upon binding to DNA breaks, PARP-1 activity increases about 500-fold over low basal levels. Studies have shown that PARP-1 and PARP-2 ADP-ribosylation activity is necessary for DNA repair processes, highlighting the vital role of this PTM in maintaining nuclear integrity (D’Amours et al., 1999; Wei and Yu, 2016).

Histone proteins are one of the biggest acceptors of ADP-ribosylation in the cell. Previous work has determined that all five histone proteins can be ADP-ribosylated in vivo (Boulikas, 1988; Huletsky et al., 1985; Leidecker et al., 2016; Messner et al., 2010; Riquelme et al., 1979; Rosenthal et al., 2015). In order to understand how histone ADP-ribosylation is involved in the DNA damage response and repair, it is important to first determine where modifications occur. However, ADP-ribosylation is a particularly challenging PTM to study for several reasons: (1) it is highly heterogeneous with a variable number ADP-ribose monomers being added to a given acceptor side chain, (2) each ADP-ribose unit imparts two negative charges to the acceptor protein, and (3) it is a relatively labile PTM. ADP-ribosylation was discovered in the 1950’s, but these challenges have precluded rigorous analysis of ADP-ribosylation sites for decades. Previous studies utilizing chemical susceptibility techniques, however, have been successful in identifying a few histone ADP-ribosylation sites in vivo (Adamietz and Rudolph, 1984; Boulikas, 1988; Bredehorst et al., 1978; Ogata et al., 1980). These sites, including H1E2 and H1E15, have been considered canonical histone ADP-ribosylation sites as they have been known for decades.

Recently, mass spectrometry (MS) has emerged as a powerful tool to identify and quantify ADP-ribosylation modifications. However, MS methods rely on a single distinct mass shift to identify PTMs. Therefore, ADP-ribosylation modifications must be modified to impart a single searchable mass shift on the acceptor residue. Previous studies have used enzymes to accomplish this task. For example, phosphodiesterase (PDE) and Nudix hydrolases have been used to digest ADP-ribosylation to a single ribose-5'-phosphate moiety (Daniels et al., 2014, 2015; Martello et al., 2016; Oka et al., 1978) (Figure 3.1). One caveat of this approach is that one of the phosphate groups is preserved, thereby reducing the charge of the peptide. This charge reduction can make identification more difficult, given that MS operates in positive mode and performs optimally on peptides in +2 charge state or higher. Poly-ADP-ribose glycohydrolase (PARG) has also been used to digest ADP-ribosylation marks to a single ADP-ribose unit (Figure 3.1). However, mono-ADP-ribose groups are difficult to analyze by MS because the ADP-ribose moiety is preferentially fragmented over the peptide backbone using CID fragmentation. Researchers have determined that the use of ETD fragmentation minimizes this internal fragmentation and can allow for localization of the PTM; however, the phosphate groups of the ADP-ribose moiety still impart negative charges to the peptide of interest, complicating MS analysis (Rosenthal et al., 2015).

Chemical derivatization has also been used to reduce the heterogeneity of ADP-ribosylation marks for MS analysis. Zhang et al. demonstrated a chemical derivatization approach utilizing hydroxylamine, which converts Asp- and Glu-linked ADP-ribosylation marks to a small hydroxamic acid derivative (Δm = 15.0109 Da) (Zhang et al., 2013) (Figure 3.2). This approach has the advantage that the negatively charged phosphate groups are removed, thereby allowing higher charge states to be reached. Furthermore, the hydroxamic acid moiety does not fragment internally, allowing the modification to be localized more readily. A major caveat of this derivatization approach, however, is that it can only derivatize ester linkages found on the side chains of modified Asp- and Glu-ribosylated residues, and therefore cannot be used to identify sites on Arg, Lys, or Ser.

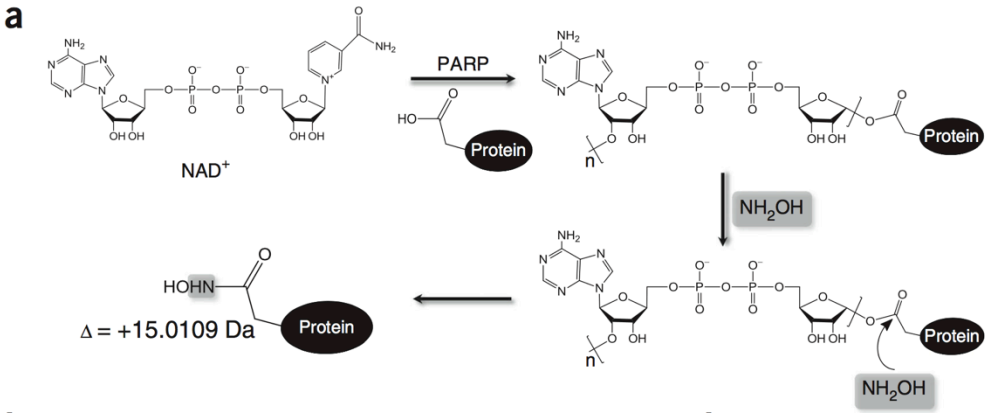


Figure 3.2. Hydroxylamine derivatization of acidic side chain acceptor sites. PARP enzymes use NAD+ as a cofactor to PARylate an acceptor protein. Hydroxylamine (NH_2OH) can be used to convert this modification to a small hydroxamic acid tag on Asp and Glu residues by attacking the ester moiety. Figure modified from (Zhang et al., 2013).

Previous studies have paved the way to identifying histone ADP-ribosylation sites. The Hottiger group used MS to identify 29 histone ADP-ribosylation sites on histones (Rosenthal et al., 2015). In this study, histones were mono-ADP-ribosylated in vitro using ARTD10, a mono-ADP-ribose transferase. More recently, Leidecker et al. identified 12 ADP-ribosylation sites on Ser residues after 10 minutes of exposure to hydrogen peroxide, a potent DNA damaging agent (Leidecker et al., 2016). Curiously, the group did not find any sites on Arg, Lys, Asp, or Glu, which have been previously shown to be modified on histone residues in vitro. These studies have provided critical insight into histone ADP-ribosylation. However, a comprehensive analysis of histone ADP-ribosylation sites in response to DNA damage is still lacking, and it remains unknown which of these histone ADP-ribosylation sites are important for DNA damage response and repair.

In this chapter, I will demonstrate work I've done to characterize the histone Asp/Glu ADP-ribosylome during DNA damage using high-resolution MS (Karch et al., 2017). We identified

30 ADP-ribosylation sites, 20 of which had not been previously discovered. We also quantified their abundances during a DNA damage time course to demonstrate that the abundance of the modification increases in a time-dependent manner and that the most accessible D/E sites are ADP-ribosylated in vivo. These results will help elucidate how histone ADP-ribosylation contributes to DNA damage repair.

3.2: Results

3.2.1: Optimization of histone digest

Histones are among the most basic proteins found in the cell and contain many Lys and Arg residues. Most proteomic experiments employ trypsin as the digestion enzyme; however, trypsin cleaves after Lys and Arg residues and therefore results in over-digestion of histone proteins. Typically, histone PTM analysis employs a derivatization approach in which propionic anhydride is used to derivatize unmodified and mono-methylated amino groups on lysine side chains and N-termini of proteins and peptides (Garcia et al., 2007). This derivatization prevents cleavage after Lys, thereby achieving an Arg-C like digestion with the efficiency of the trypsin enzyme. This procedure allows for the generation of longer, more hydrophobic peptides as cleavage after lysine is prevented and the propionyl groups impart a greater hydrophobicity onto the peptide.

Derivatization with propionic anhydride cannot be used in studies aiming to identify ADP-ribosylation sites, however. The basicity required for the derivatization causes the labile ADP-ribosylation sites to degrade (Hassa et al., 2006). The reaction cannot be performed after incubation with hydroxylamine because the hydroxamic acid tag reacts with propionic anhydride (data not shown).

Therefore, we aimed to optimize histone sequence coverage using different digestion enzymes to maximize coverage of as many potential ADP-ribosylation sites as possible. We tested three digestion enzymes: (1) chymotrypsin, which cleaves C-terminal to Phe, Trp, and Tyr, (2) pepsin, which cleaves non-specifically (but prefers hydrophobic residues), and (3) trypsin (limited digest). Limiting the reaction time of trypsin will allow for many missed cleavage events, thereby generating longer peptides that can be retained on C18 columns. Previous studies indicated that H2B and H1 are the main histone ADP-ribosylation acceptors, and so coverage of these proteins was prioritized (Boulikas, 1988).

To optimize the performance of each enzyme, we digested a histone sample under several different conditions. To preserve native PTM profiles, which can affect digestion patterns, we used histones that were extracted from HeLa cells for this analysis. For trypsin experiments, we digested samples for 30, 60, and 90 minutes. For pepsin experiments, we digested histones for 30 minutes, 2.5 hours, 6 hours, and overnight. For chymotrypsin, we digested histones at two different enzyme:substrate ratios, 1:20 and 1:100, which can alter enzyme specificity. The resulting digests were analyzed on our Thermo Q Exactive instrument.

The results of the digestion optimization experiment are shown in Figure 3.3. The percent coverage is displayed as an average for all variants detected for that histone. Panel D shows the condition with the best coverage for each enzyme tested. The trypsin digest (1 hour) had the highest coverage of all histones, including H1 and H2B, and yielded the highest number of potential modification sites (Table 3.1). We therefore chose to move forward with the 1 hour limited trypsin digest. One caveat of a limited digestion approach is that the data analysis is more complex due a larger number of missed cleavage events and consequently a more varied pool of peptides.

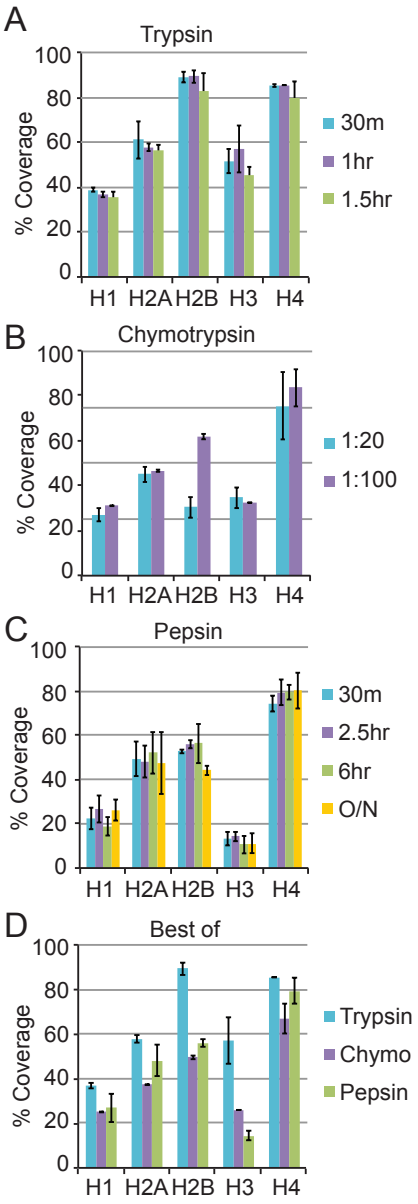


Figure 3.3. Optimization of histone coverage. Histones extracted from HeLa cells were digested by proteases under several different experimental conditions to optimize coverage. The coverage obtained for each histone type are shown for trypsin (A), chymotrypsin (B), and pepsin (C). Values represent an average coverage value for all detected histone variants. The color of each bar indicates the experimental condition as shown in the key. Panel D shows the experimental condition with the highest coverage for each protease tested. Error bars represent

the standard deviation of three experimental replicates. Figure and caption taken from (Karch et al., 2017).

Table 3.1. Histone coverage optimization.

Histone		Pepsin	Chymotrypsin	Trypsin
H1	Coverage	26.53%	24.90%	36.57%
	# missing	2	4	0
	# included	5	3	7
H2A	Coverage	48.03%	37.19%	57.66%
	# missing	4	4	1
	# included	5	5	8
H2B	Coverage	56.01%	49.41%	89.33%
	# missing	4	4	1
	# included	6	6	9
H3	Coverage	14.21%	25.74%	56.77%
	# missing	8	7	6
	# included	3	4	5
H4	Coverage	79.29%	66.99%	85.44%
	# missing	0	0	0
	# included	7	7	7
Total	# missing	18	19	8
	# included	26	25	36

Rows labeled “# missing” and “# included” indicate the number of Asp/Glu sites that were not covered or included in the identified peptides, respectively, from the isoform with the highest coverage. The coverage values are given as an average of the identified histone variants. Table taken from (Karch et al., 2017).

3.2.2: Identification of histone ADP-ribosylation sites catalyzed by PARP-1 in vitro

Our next objective was to optimize sample preparation and MS procedures to identify as many histone ADP-ribosylation sites as possible. Given the negative charge, propensity to degrade, and heterogeneous nature of ADP-ribosylation modifications, we chose to derivatize samples with hydroxylamine prior to MS analysis. This procedure enables identification of Asp and Glu ADP-ribosylation sites by leaving a small hydroxamic acid tag on the side chain (Δm = 15.0109 Da), but precludes analysis of Lys, Arg, and Ser sites (Zhang et al., 2013).

We wanted to optimize this experimental procedure using histones that are ADP-ribosylated in vitro. To this end, we developed an in vitro assay using PARP-1, which has been shown to be critical for DNA damage response, and its ADP-ribosylation activity is required for repairing DNA lesions (D'Amours et al., 1999; Wei and Yu, 2016). PARP-1 has also been shown to modify all five histone proteins in vitro (Boulikas, 1988; Huletsky et al., 1985; Leidecker et al., 2016; Messner et al., 2010; Riquelme et al., 1979; Rosenthal et al., 2015). Given that PTM profiles likely affect the locations and abundances of ADP-ribosylation, we opted to use histones that were extracted from HeLa cells for this assay so that native PTM profiles are present. In the assay, histones were combined with PARP-1, NAD+ to serve as an ADP-ribose donor, and a small double stranded DNA segment to serve as an activator for PARP-1 activity.

We confirmed that the assay works by performing a Western blot against poly-ADP-ribose (PAR) (Figure 3.4A). The results demonstrate that no ADP-ribosylation occurs in the absence of NAD+ as expected, but a robust signal is detected when NAD+ is included in the reaction. The resulting "smear" on the Western blot is typical of ADP-ribosylation modifications due to the heterogeneity of the modification. A variable number of ADP-ribose units can be added to a given modification site, and a variable number of sites can be modified, thus allowing the modified proteins to adopt a wide range of molecular weights (Boulikas, 1988).

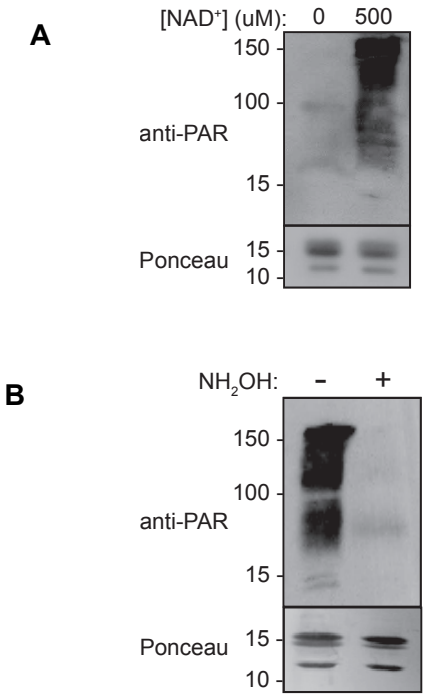


Figure 3.4. PARP-1 ADP-ribosylates Asp and Glu residues of histone proteins in vitro. (A) In vitro PARP-1 assay effectively ADP-ribosylates histone proteins. Total histones extracted from HeLa cells (20 ug) were incubated with PARP-1 and a short double-stranded DNA molecule with or without NAD+ for 30 minutes at 30˚C. The reaction was quenched by freezing and a Western blot against poly-ADP-ribose was performed. Prior to transfer, membranes were stained with Ponceau to verify equal loading between lanes. (B) Histone ADP-ribosylation occurs predominantly on Asp/Glu residues. The same in vitro PARP-1 assay was performed on 40 ug of histones extracted from HeLa cells. Half of the sample was incubated with 1 M hydroxylamine overnight at room temperature. A Western blot against polyADP-ribose was performed to monitor incorporation of ADP-ribosylation. Figure and caption taken from (Karch et al., 2017).

Given that previous studies have found ADP-ribosylation sites on Lys, Arg, and Ser residues, which this method cannot address, we sought to determine the how much ADP-

ribosylation occurs on Glu/Asp compared to Lys/Arg/Ser. This analysis would give us an idea of how many sites may be precluded from analysis. To accomplish this task, we performed an in vitro ADP-ribosylation assay and treated half of the sample with hydroxylamine overnight. However, we utilized biotin-NAD+, which can still be used in catalysis by PARP-1, in place of unlabeled NAD+. Therefore, we can use streptavidin-HRP as a detection reagent in place of the PAR antibody. The streptavidin-HRP reagent provides greater sensitivity given that it can detect as low as a single ADP-ribose unit, while the PAR antibody can only detect polymers of ADP-ribose. We can therefore detect differences in susceptibility to hydroxylamine of both MARylation and PARylation.

Derivatization with hydroxylamine will cause the biotin tag to be removed from Asp and Glu residues, causing the signal to be reduced, while hydroxylamine-insensitive sites (Arg, Lys, Ser) will retain the biotin tag, causing the signal to be retained. Treatment with hydroxylamine caused a near complete depletion of streptavidin-HRP signal, indicating that a vast majority of histone ADP-ribosylation sites occur on Glu and Asp residues compared to Lys, Arg, and Ser (Figure 3.4B).

We performed an additional control to ensure that hydroxylamine cannot chemically modify Glu and Asp side chains, to ensure that no false positive identifications are made. To this end, we incubated recombinant histone proteins with hydroxylamine overnight as done in the standard protocol. The use of recombinant, unmodified histones in place of histones extracted from HeLa cells ensures that we do not identify any endogenous sites. We did not identify any hydroxamic acid modifications on Glu or Asp, indicating that there is no chemical modification occurring and that hydroxylamine is a suitable reagent to use for derivatization of the ADP-ribosylation modification.

We next digested the hydroxylamine derivatized samples from the PARP-1 ADP-ribosylation assay and analyzed them by high-resolution MS/MS on our Thermo Orbitrap Fusion instrument. We created a method that performs HCD fragmentation on peptides with charge

states +2 to +4 or ETD fragmentation on peptides with a charge state of +4 or higher. Generally, higher charged peptides have better fragmentation with ETD, and therefore this method aims to confidently identify and localize hydroxamic acid-containing peptides of all charge states.

We were able to identify 7 modification sites in vitro using this assay (Table 3.2). However, well-characterized ADP-ribosylation sites, such as H1E2 and H1E15, were not detected in this experiment (Ogata et al., 1980). Furthermore, we did not detect any ADP-ribosylation modifications on H2A or H3, despite the fact that it has been determined that PARP-1 can modify all five histones.

Table 3.2. Identified ADP-ribosylation sites. The modified residue is given in bold red. (Ac) indicates protein N-terminal acetylation. In columns "in vivo" and "in vitro," the Y (standing for "Yes") designation indicates that the peptide was found in the respective experiments. The column labeled "z" represents the charge of the peptide. The column labeled "Lit" (standing for "literature") indicates in which previous studies, if any, identified that modification (1 = Ogata et al., 1980, 2 = Rosenthal et al., 2015, 3 = Zhang et al., 2013).

#	Histone	Site	Sequence	Peptide m/z	z	Parent Mass	Score	Accession	in vivo	in vitro	Lit
1	H1.1	E55	**E**RGGVSLAALK	558.331	2	1115.655	4.51E-04	Q02539	Y		
2	H1.2	E2	S(Ac)**E**TAPAAPAAAPPAEK	768.389	2	1535.771	4.60E-05	P16403	Y		1
			S(Ac)**E**TAPAAPAAAPPAEKAPVK	966.015	2	1931.023	1.02E-03	P16403			
3	H1.2	E15	S(Ac)ETAPAAPAAAPPA**E**K	768.389	2	1535.770	8.68E-05	P16403	Y		1
4	H1.2	E41	ASGPPVS**E**LITK	607.343	2	1213.679	6.22E-04	P16403	Y	Y	
			KASGPPVS**E**LITK	671.390	2	1341.773	6.76E-05	P16403			
			KASGPPVS**E**LITK	447.930	3	1341.773	4.24E-04	P16403			
5	H1.2	E52	**E**RSGVSLAALK	573.337	2	1145.667	4.29E-04	P16403	Y		
6	H1.2	D71	ALAAAGY**D**VEK	561.794	2	1122.581	1.79E-04	P16403	Y		
7	H1.2	E73	ALAAAGYDV**E**K	561.794	2	1122.581	2.90E-04	P16403	Y		
			KALAAAGYDV**E**K	625.840	2	1250.673	3.81E-05	P16403			
8	H1.4	E2	S**E**TAPAAPAAPAPAEK	747.387	2	1493.766	2.98E-03	P10412	Y		
9	H1.5	D74	ALAAGGY**D**VEKNNSR	790.392	2	1579.776	9.80E-03	P16401	Y		
			KALAAGGY**D**VEKNNSR	569.960	3	1707.865	1.73E-01	P16401			

10	H1.5	E76	ALAAGGYDVEK	554.784	2	1108.560	2.00E-02	P16401	Y		
			ALAAGGYDVEKNNSR	790.390	2	1579.773	7.96E-02	P16401			
			KALAAGGYDVEKNNSR	569.961	3	1707.866	3.64E-02	P16401			
11	H1.5	D74 & E76	ALAAGGYDVEKNNSR	797.894	2	1594.780	5.83E-02	P16401	Y		
			ALAAGGYDVEK	562.285	2	1123.562	3.62E-01	P16401			
12	H2A2B	D90	HLQLAVRNDEELNK	565.300	3	1693.885	7.12E-04	Q8IUE6	Y		
13	H2A2B	D90 & E91	HLQLAVRNDEELNK	854.953	2	1708.899	1.52E-02	Q8IUE6	Y		2
14	H2AZ	D93	GDEELDSLIK	567.287	2	1133.567	3.30E-02	P0C0S5	Y		
15	H2B1C	E35	KESYSVYVYK	640.832	2	1280.656	1.87E-02	P62807	Y		2
16	H2B1C	D51	QVHPDTGISSK	395.208	3	1183.607	8.13E-03	P62807	Y		
17	H2B1C	E93	EIQTAVR	416.235	2	831.462	4.11E-01	P62807	Y		
18	H2B1C	E105	LLLPGELAK	484.811	2	968.614	4.39E-03	P62807	Y	Y	2
19	H2B1C	E113	HAVSEGTKAVTKYTSSK	603.655	3	1808.949	2.16E-02	P62807	Y		
20	H2B3B	E93	STITSREVQTAVR	731.898	2	1462.788	8.77E-03	Q8N257	Y		
21	H3.1	E59	STELLIR	423.756	2	846.505	3.26E-03	P68431		Y	
22	H3.1	E73	EIAQDFKTDLR	450.906	3	1350.701	2.07E-02	P68431	Y	Y	
23	H3.1	D77	EIAQDFK	433.224	2	865.441	4.49E-02	P68431	Y		2
			EIAQDFKTDLR	675.849	2	1350.691	3.82E-01	P68431			
			EIAQDFKTDLR	450.902	3	1350.690	1.94E-01	P68431			
24	H3.1	D81	EIAQDFKTDLR	675.858	2	1350.709	2.99E-01	P68431	Y		
25	H4	D24	DNIQGITKPAIR	447.591	3	1340.758	1.20E-03	P62805	Y		
26	H4	E52	ISGLIYEETR	598.321	2	1195.634	4.06E-03	P62805	Y	Y	2
27	H4	E53	ISGLIYEETR	598.321	2	1195.634	1.64E-03	P62805	Y	Y	2
28	H4	E63	VFLENVIR	502.798	2	1004.589	1.32E-02	P62805	Y		2
29	H4	D68	DAVTYTEHAK	575.281	2	1149.554	2.32E-03	P62805	Y		
30	H4	E74	DAVTYTEHAK	575.281	2	1149.555	1.67E-03	P62805	Y	Y	2,3

Table was modified from (Karch et al., 2017).

There are several possible explanations of why more sites were not identified. Since this was an in vitro study utilizing only PARP-1, we did not have the full complement of ART enzymes. Some sites may be mediated by ART enzymes other than PARP-1 and were therefore not

modified in this experiment. Secondly, PARP-1 may require some effector proteins to modify its full complement of acceptor sites that were not present in this study. Thirdly, unfolded histone substrates were used in the assay. Folded substrates with their native PTM profiles may be required to achieve full PARP-1 function. And lastly, the reaction may not be efficient and so the modifications may be too low in abundance to be identified by MS. Nonetheless, these in vitro studies demonstrate the utility of hydroxylamine derivatization for ADP-ribosylation detection and enabled us to develop an MS platform for analysis of hydroxamic acid-containing peptides.

3.2.3: In nucleo incubation with NAD+ leads to spurious ADP-ribosylation of histone proteins

We next sought to identify more physiologically relevant ADP-ribosylation sites by performing experiments in HeLa cells. However, given that ADP-ribosylation is a low abundance modification compared to more canonical PTMs such as acetylation, we aimed to increase the abundance of the modification in nucleo to facilitate identification (D'Amours et al., 1999). To this end, we treated HeLa cells with dimethyl sulfate (DMS), a potent DNA damaging agent, isolated the nuclei, and incubated the nuclei with additional NAD+ (Figure 3.5). NAD+ is impermeable to the outer membrane of cells but can cross the nuclear membrane. Therefore, incubation of nuclei with NAD+ enables more ADP-ribosylation to occur in the nucleus. These in nucleo experiments are common in the chromatin biology field, especially in studies utilizing radiolabeled NAD+ as it enables the radiolabeled NAD+ to easily enter the nucleus (Adamietz and Rudolph, 1984; Boulikas, 1988; Riquelme et al., 1979).

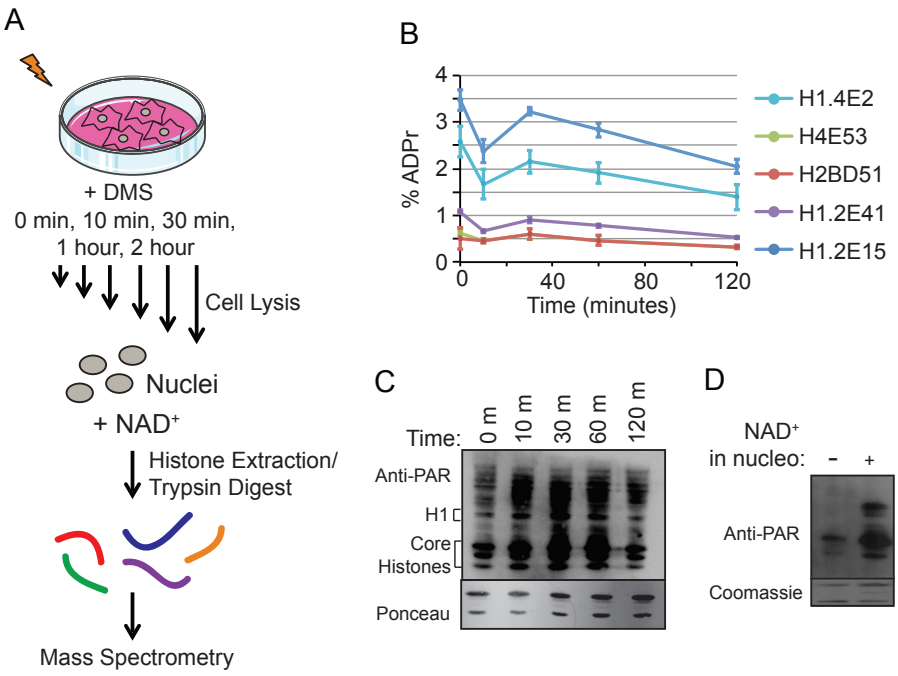


Figure 3.5. Incubation of HeLa nuclei with NAD+ leads to spurious histone ADP-ribosylation. (A) General workflow used to identify histone ADP-ribosylation sites. HeLa cells were treated with 0.2 mM DMS for indicated time points, hypotonically lysed, and incubated with 0.2 mM NAD+ for 2 hours at 37˚C. (B) Histone ADP-ribosylation levels remain constant over the DMS timecourse. Five example ADP-ribosylation sites were quantified and normalized to the abundance of the non-modified peptide. Error bars represent the standard deviation of three experimental replicates. (C) Histones are highly ADP-ribosylated upon incubation with NAD+ in nucleo. A Western blot against poly-ADP-ribose was performed on histones extracted from HeLa cells during the DMS timecourse. (D) Histones undergo a larger degree of ADP-ribosylation when incubated with NAD+ in nucleo compared to histones extracted from intact cells. A Western blot against poly-ADP-ribose was performed on histones extracted from HeLa cells with or without in nucleo incubation with NAD+. Figure and caption taken from (Karch et al., 2017).

DMS induces DNA damage by alkylating purine bases, which are then repaired using the base excision repair (BER) pathway (Cabelof et al., 2002). Given that ART enzymes are dramatically activated during DNA damage and repair, treatment of the HeLa cells with DMS will increase the abundance of ADP-ribosylation on histone proteins. We treated cells with 0.2mM DMS for 0, 10, 30, 60, and 120 minutes. Immediately following treatment, we extracted the nuclei of the cells using hypotonic lysis and incubated them with 0.2mM NAD+ at 37˚C for 2 hours. Then, we performed an acid extraction to isolate histone proteins.

We then performed a Western blot against PAR to determine the degree of PARylation in the samples (Figure 3.5C). The results show a high amount of PARylation across all time points, even the untreated control (Fig 3.5B and 3.5C). This result was unexpected because it has been reported that basal levels of ADP-ribosylation are very low. From there, we digested the histone samples with trypsin and performed high resolution MS/MS analysis and identified 16 modification sites. We quantified 5 of these sites across time points and found that the abundances were consistently high (Fig 3.5B). We expected to observe an increase in ADP-ribosylation abundance over time as DNA damage accumulates over time during DMS treatment. We believe that this unexpectedly high abundance of ADP-ribosylation can be attributed to the incubation of NAD+ in nucleo. Flooding the cell with substrate (NAD+) can lead to excessive ART activity well beyond basal levels regardless of the level of DNA damage.

We next sought to determine how much ADP-ribosylation can be attributed to the in nucleo NAD+ incubation relative to normal levels of ADP-ribosylation. To this end, we harvested untreated HeLa cells and subjected them to the in nucleo incubation with NAD+ and performed a Western blot against PAR (Figure 3.5D). We found that the level of ADP-ribosylation in cells that underwent incubation with NAD+ was dramatically higher than the same cells where the incubation step was omitted. Therefore, although the incubation step increased the abundance of ADP-ribosylation to facilitate detection by MS, it should not be used for quantification as it artificially increases the abundance of ADP-ribosylation sites in vivo. Furthermore, it may lead to

modification of sites that would otherwise remain unmodified under normal cellular levels of NAD+.

3.2.4: Boronate enrichment enhances identification of ADP-ribosylation sites

We next sought an alternative strategy to enrich ADP-ribosylation modifications given that the in nucleo incubation with NAD+ lead to spurious ADP-ribosylation. We utilized boronate chromatography to achieve this goal, as has been previously cited in the literature (Zhang et al., 2013). Boronate covalently binds cis-diols, 2 of which are present in the ADP-ribose modification (Figure 3.6A). Due to the covalent nature of ADP-ribose interaction with boronate, stringent washes can be used to remove peptides that are not ADP-ribosylated. Incubation of the boronate column with hydroxylamine overnight allows for simultaneous derivatization and elution of modified peptides from the column. Eluted peptides can then be identified and quantified using high-resolution MS/MS.

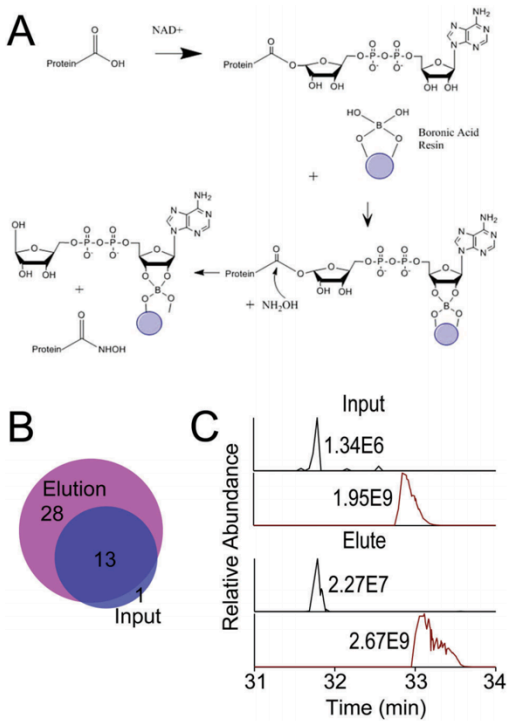


Figure 3.6. Boronate enrichment enables identification of low-level ADP-ribosylation sites on histones. (A) Mechanism of covalent interaction between boronate and cis-diols of an ADP-ribose modification. Hydroxylamine removes ADP-ribose modifications on acidic residues by nucleophilic attack at the carbonyl group of the resulting ester bond. (B) Comparison of sites identified in input and enriched samples. (C) Boronate enrichment increases the abundance of modified peptides relative to unmodified peptides. Extracted ion chromatograms (XICs) are displayed for the peptide (sequence: KASGPPVSELITK, H1.2) in its modified (hydroxamic acid, top trace, black, $[M + 2H]^{2+}$ = 671.301 m/z) and unmodified (bottom trace, maroon, $[M + 2H]^{2+}$ = 663.885 m/z) forms. The top figure represents XICs from input sample and the bottom figure represents XICs from the elution sample. The peptide shown here underwent a 12.4-fold enrichment.

We aimed to the test the utility of boronate enrichment for analysis of histone ADP-ribosylation using histones extracted from untreated HeLa cells. As mentioned previously, under normal physiological conditions, levels of histone ADP-ribosylation are very low. Basal ADP-ribosylation activity is necessary for some nuclear processes such as transcription and regulation of chromatin structure (Martinez-Zamudio and Ha, 2012; Murcia et al., 1986; Niedergang et al., 1985). We decided to use untreated samples because very low-level ADP-ribosylation sites will be the most difficult ones to detect.

We extracted histones from a HeLa cell pellet that was harvested under standard growing conditions. We then digested the histone proteins and enriched ADP-ribosylated peptides using boronate chromatography. We reserved a portion of the input sample to use as a control. Samples were eluted and derivatized on-column by incubation with 1M hydroxylamine overnight. Additional 1M hydroxylamine was pushed through the column the next day to ensure maximal elution from the column.

The results indicate that the boronate enrichment did allow for a larger number of modified peptides to be identified in the elution compared to the input sample (28 sites compared to 14, Figure 3.6B). To illustrate this, we show the chromatograms of an example peptide from H1.2 in input and enriched samples (sequence: KASGPPVSELITK, unmodified $[M + 2H]^{2+}$ = 663.885 m/z, modified $[M + 2H]^{2+}$ = 671.301 m/z) (Figure 3.6C). We determined that this peptide experienced a 12.4-fold enrichment after enrichment with boronate, which enabled it to be detected by MS. These results demonstrate the utility of boronate chromatography to facilitate identification of ADP-ribosylation sites with MS. However, quantification must be completed in the input sample because the abundance of the unmodified peptide is needed for label-free relative abundance calculations.

3.2.5: Identification and quantification of histone ADP-ribosylation levels during DMS-induced DNA damage time course

We next sought to comprehensively analyze and quantify histone ADP-ribosylation sites in vivo during a DNA damage time course to understand which sites are important for DNA damage detection and repair. To this end, we treated HeLa cells with 0.2mM DMS, a DNA damaging agent, for 0, 10, 30, 60, and 120 minutes. We first validated that this dosage was adequate to initiate the DNA damage response (DDR) by monitoring the level of of H2AX phosphorylation at Ser 139 (γ-H2AX) over time by Western blot (Figure 3.7A). Previous work has shown that γ-H2AX foci are formed rapidly, within seconds, of DNA damage in the cell and so detection of γ-H2AX is a standard technique to determine if DDR is occurring (Sharma et al., 2012). The results demonstrate that the level of γ-H2AX increases over time, indicating that the cells are undergoing DDR and that this DMS dosage is adequate for determining which ADP-ribosylation sites are involved in DNA damage repair processes. Furthermore, γ-H2AX levels are very low in the untreated control as expected.

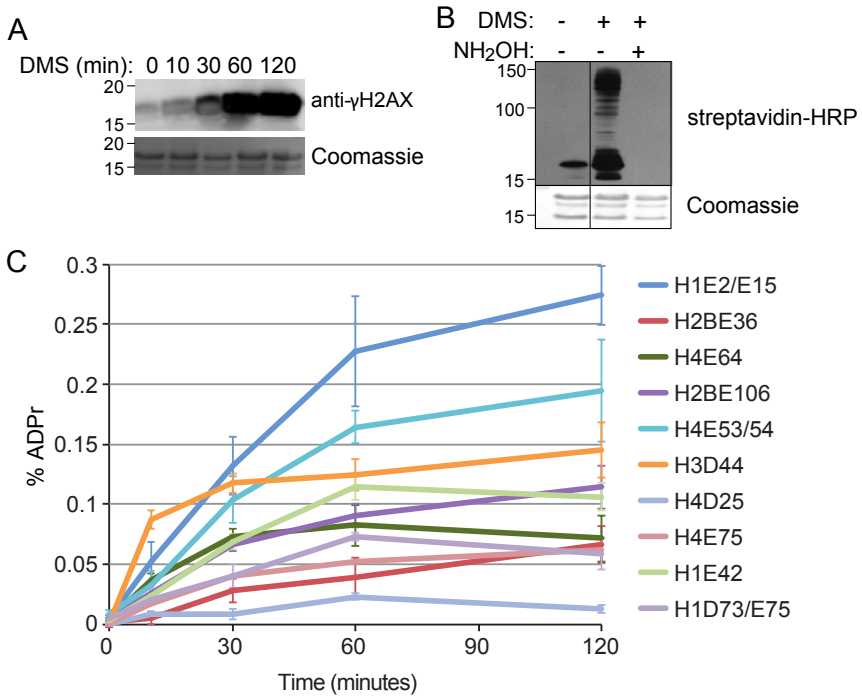


Figure 3.7. Histone ADP-ribosylation levels increase with the amount of DNA damage. (A) DMS treatment leads to an extensive DNA damage response. Histones were extracted from cells undergoing DNA damage by DMS treatment. A Western blot against H2AX phosphorylated at Ser139 (γ-H2AX) was performed. (B) Most histone ADP-ribosylation in response to DMS damage occurs on Asp/Glu residues. Cells were treated with 0.2 mM DMS for 2 hours, nuclei were extracted and incubated with biotin–NAD+. A portion of the sample was incubated with 1 M hydroxylamine overnight at room temperature with shaking. A Western blot using streptavidin-HRP was used to visualize the extent of ADP-ribosylation. Lanes 2 and 3 of the blot were removed because they were not relevant. The blot and Coomassie stains show lanes 1, 4, and 5 of the original Western blot. (C) Histone ADP-ribosylation levels increase throughout DNA damage timecourse. Ten ADP-ribosylation sites were quantified and plotted on the graph. Error bars represent the standard error of 5 biological replicates. Label-free quantification was performed by normalizing the abundance of the peptide containing an ADP-ribosylation site to the

abundance of the peptide in all of its modified and unmodified forms. Figure and caption taken from (Karch et al., 2017).

Given that the hydroxylamine derivatization method used here cannot identify ADP-ribosylation sites on Lys, Arg, or Ser, we sought to determine how many of the ADP-ribosylation sites on histones in vivo can be attributed to these residues compared to Glu and Asp, which we can analyze. To this end, we treated HeLa cells with 0.2mM DMS for 2 hours, harvested the cells, and isolated the nuclei. We then incubated the nuclei with biotin-labeled NAD+ for 2 hours at 37˚C to allow modification sites to become biotinylated. We then treated half of the sample with hydroxylamine overnight and performed a Western blot with streptavidin-HRP to determine the levels of ADP-ribosylation before and after derivatization with hydroxylamine. The use of biotinylated-NAD+ allows for greater sensitivity than PAR Western blots because it can detect a single biotinylated ADP-ribose unit where PAR antibodies can detect chains of ADP-ribose. As described in Chapter 3.2.4, in nucleo incubation with NAD+ leads to increased levels of ADP-ribosylation, so we also included an untreated control that was incubated with biotin-NAD+ to determine the level of background labeling of biotinylated NAD+ in untreated cells.

The results demonstrate that the DMS treatment lead to a significant amount of histone ADP-ribosylation (Figure 3.7B, lane 2). Overnight treatment with hydroxylamine removed nearly all of the ADP-ribosylation modification (lane 3), as demonstrated by the very low signal on the Western blot. Notably, the hydroxylamine-treated sample (lane 3) has even lower signal than the untreated control sample (lane 1), indicating that nearly all detectable ADP-ribosylation sites are present on Glu/Asp rather than Lys/Arg/Ser. Therefore, the methods used here can detect most histone ADP-ribosylation sites in these conditions.

We next sought to identify which histone ADP-ribosylation sites are involved in the DNA damage response and subsequent repair by base excision repair (BER) pathways. To

accomplish this task, we treated cells with DMS, an alkylating agent that induces single strand DNA breaks through the BER pathway, for 0, 10, 30, 60, and 120 minutes. We then harvested cells, extracted histones, digested them with trypsin, and derivatized them overnight with hydroxylamine. We then performed high-resolution MS/MS analysis on our Thermo Fusion instrument and were able to identify 29 ADP-ribosylation sites, 10 of which had high enough abundance to accurately quantify over the time course (Figure 3.7C). We quantified sites with label-free relative quantification. To this end, we summed the intensities of each peptide containing the modification of interest and dividing it by the total intensity of the peptides containing that modification site in all of their modified forms and charge states. The results demonstrate that each ADP-ribosylation site increases in abundance over the time course. This result is expected because DMS treatment will cause an accumulation of DNA damage over time, consequently increasing the degree of DNA damage response and repair processes, which include ADP-ribosylation of histone proteins. Figure 3.8 highlights the histone residues that were found to be modified in this analysis. Note that not all sites are included because the structure does not contain H1, and portions of the histones, particularly N-terminal tails, are not present in the structure.

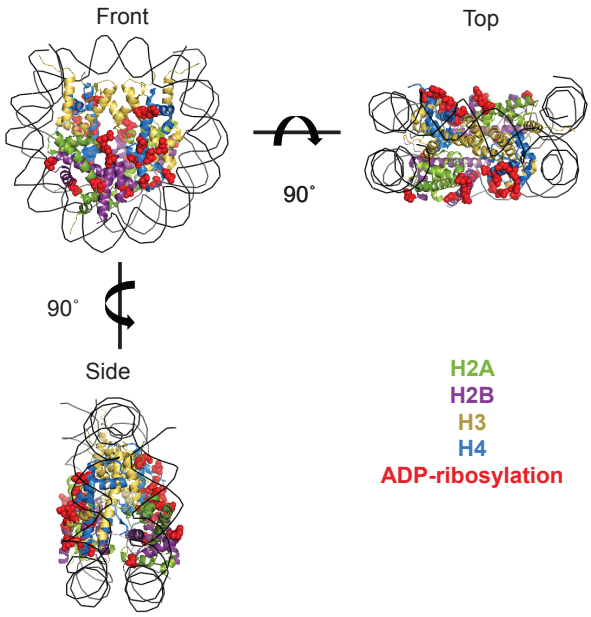


Figure 3.8. ADP-ribosylation occurs primarily on the surface of the nucleosome. The crystal structure of the human nucleosome containing H3.1, H4, H2A type 1-B/E, and H2B type 1-J (PDB: 3W98) is shown. The histones are colored according to the legend, with Asp/Glu residues that were found to be ADP-ribosylated shown in red spheres. Note that PDB structure does not contain full length histones and so not all identified sites are displayed in the figure (structure contains H3.1: AA 38-136, H4: 19-104, H2A type 1-B/E: 13-118, H2B type 1-J: 30-124). Figure and caption taken from (Karch et al., 2017).

3.3: Discussion

ADP-ribose transferase (ART) enzymes are critical for DNA damage repair in cells. PARP-1 and PARP-2 directly bind single and double strand DNA breaks to ADP-ribosylate target proteins, including themselves (Langelier et al., 2012). This activity acts as the first step to recruit DNA damage repair proteins to the site of damage. ADP-ribosylation by ART enzymes has been shown to be critical for nearly all types of DNA repair pathways including homologous

recombination (HR), non-homologous end-joining (NHEJ), and alternative-NHEJ (alt-NHEJ) (Pears et al., 2012).

Despite the great advances in understanding the role of ARTs in DNA damage repair pathways, the role of ADP-ribosylated acceptor proteins in these pathways is not well known. However, several studies have aimed to identify ADP-ribose acceptor proteins upon different types of DNA damage and cellular stress as a first step towards this aim (Daniels et al., 2014; Martello et al., 2016; Zhang et al., 2013). For example, Daniels et al. used phosphodiesterase (PDE) to digest ADP-ribosylation modifications to a single ribose-5'-phosphate moiety (Daniels et al., 2014). The researchers then enriched modified peptides using phospho-enrichment strategies and analyzed them by MS. These types of studies have identified histones as a major acceptor of ADP-ribosylation. However, it is still unknown how ADP-ribosylated histones are involved in the DNA damage response and DNA damage repair.

The first step towards understanding the role of histone ADP-ribosylation in these processes is determining where these modifications occur. There have been several studies aiming to achieve this goal, many of which have been performed in vitro on recombinant histone substrates. However, Leidecker et al. analyzed histone ADP-ribosylation sites in human osteosarcoma cells after treatment with hydrogen peroxide, a DNA damaging agent (Leidecker et al., 2016). They used a Nudix hydrolase to digest the ADP-ribosylation modifications to phospho-ribose and identified the sites using MS. They identified 12 novel ADP-ribosylation sites, all located on serine residues. Previous studies have shown that Asp and Glu are also major acceptors of ADP-ribosylation on histones, illuminating the need for a comprehensive analysis of the Asp/Glu ADP-ribosylome on histone proteins.

Here, we demonstrate the first comprehensive analysis of the histone Asp/Glu ADP-ribosylome during DNA damage and repair (Karch et al., 2017). We are also able to provide the first quantification of histone ADP-ribosylation levels during these processes. Overall, we were able to identify 30 histone ADP-ribosylation sites, 20 of which are novel. It is important to note

that most of the identified modification sites are located on the surface of the nucleosome (Figure 3.8). In fact, nearly every exposed Asp/Glu residue of the nucleosome was found to be ADP-ribosylated in this study. Furthermore, there are several Asp/Glu residues located in the interior of the nucleosome, not exposed to the surface, that were not identified in this study or in previous publications. This finding implies that ART enzymes responding to DNA damage are ADP-ribosylating nearly any exposed acceptor residue on nearby nucleosomes.

Previous studies have shown that ADP-ribosylation of nucleosomes destabilizes histone-DNA interactions and increases accessibility to DNA (Martinez-Zamudio and Ha, 2012). It has been hypothesized that this destabilization is due to charge repulsion between the negatively charged ADP-ribose modifications and the negatively charged DNA contained in the nucleosome. Indeed, PARP-1 interaction with DNA is mediated in this fashion (Steffen et al., 2016). PARP-1 binding to DNA lesions activates its ADP-ribosylation levels about 500-fold over low basal levels. This activation leads to automodification of PARP-1 itself as well as histone proteins and other acceptors. This automodification has been shown to be involved in relaxation of chromatin structure, recruitment of DNA damage repair proteins, and ultimately, dissociation of PARP-1 from chromatin (Ferro and Olivera, 1982; Hassa et al., 2006; Mortusewicz et al., 2007; Strickfaden et al., 2016).

The discovery that nearly every exposed Asp/Glu residue of the nucleosome can is ADP-ribosylated during DNA damage detection and repair, it seems unlikely that modification of specific residues is critical for mediating these processes. Rather, it implies that ART enzymes are acting through a “brute force” mechanism in which extensive ADP-ribosylation of the nucleosomal surface enables relaxation of chromatin and access of the underlying damaged DNA to allow repair factors to interact with the lesion(s).

3.4: Methods

3.4.1: Cell culture and histone extraction

HeLa S3 cells were grown on 10cm plates in DMEM media supplemented with 10% newborn calf serum (Gibco), penicillin-streptomycin solution (Fisher, 10,000 units penicillin, 10 mg/mL streptomycin) and 1X GlutaMAX (Fisher). For DNA damage experiments, cells were treated at the indicated concentration by addition of dimethylsulfate (DMS; Sigma) by addition to the media. Cells were treated when they reached approximately 70% confluency. To harvest cells, media was removed and the cells were washed with 10 mL sterile PBS. Cells were removed by scraping in PBS and subsequently pelleted at 1,000 rpm for 4 minutes. PBS was removed and cells were either flash frozen to be saved for histone extraction, or, for the in nucleo experiments, nuclei were immediately extracted using hypotonic lysis. To lyse cells, cell pellets were resuspended in 5X volume of ice cold buffer (1M HEPES, pH 7.9, 1M $MgCl_2$, 2.5M KCl, 1M DTT, 1X Halt protease in dH_2O). Pellets were incubated on ice for 5 minutes and dounced homogenized (30 strokes). Cells were pelleted at 600g for 5 minutes at 4˚C. Supernatant was then removed and nucei pellet was resuspended and incubated in buffer (15mM Tris-HCl, pH7.5, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 1mM $CaCl_2$, 250mM sucrose, 0.2mM NAD+) at 37˚C for two hours on a rotator. After incubation, cells were pelleted at 600g for 5 minutes. The supernatant was removed and histones were extracted using a standard sulfuric acid extraction followed by TCA precipitation as previously described (Karch et al., 2016; Lin and Garcia, 2012; Sidoli et al., 2016). For experiments where the in nucleo incubation with NAD+ was not performed, nuclei were extracted from cell pellets with detergent as previously described (Karch et al., 2016). The same histone extraction was performed as described above.

3.4.2: In vitro ADP-ribosylation Assay

Full-length human PARP-1 (residues 1-1014) with an N-terminal histidine tag was expressed in E. coli strain BL21(DE3) Rosetta2 (Novagen). PARP-1 was chromatographically purified using Ni2+-affinity, heparin, and gel filtration as previously described (Langelier et al., 2017).

All in vitro ADP-ribosylation assays were performed in 25uL of buffer (50mM Tris-HCl, pH 8.0, 4mM MgCl2, 250uM DTT, and 20mM NaCl). Each reaction contained 10pmol of recombinant wild-type PARP-1, 200uM NAD+, 20ug histones extracted from HeLa cells, and 10pmol of a synthesized double-stranded duplex DNA that was synthesized as previously described (Langelier et al., 2017). Reactions were incubated at 37˚C for 30 minutes and quenched by flash freezing.

3.4.3: Sample Digestion

untreated HeLa cells. Protein concentrations were measured using a Bradford assay.

In trypsin experiments, histones were resuspended in ammonium bicarbonate, pH 8.0 to a concentration of 0.1 ug/uL. Trypsin was added in a 1:20 enzyme:substrate ratio and incubated for 30 minutes, 1 hour, or 1.5 hours at 37˚C. Reactions were quenched by addition of glacial acetic acid to lower the pH to approximately 4 followed by freezing at -80˚C.

In chymotrypsin experiments, histones were resuspended in ammonium bicarbonate, pH 8.0, to a concentration of 0.1 ug/uL. Chymotrypsin was added in a 1:20 or 1:100 enzyme:substrate ratio and incubated at 37˚C for 6 hours and subsequently quenched by freezing.

3.4.4: Derivatization and desalting

In cases where samples were derivatized, hydroxylamine (Sigma) was added to a final concentration of 1M immediately after histone extraction to minimize loss of the modification over time. Samples were incubated overnight at room temperature with shaking. Samples were desalted using home-made stage tip columns as previously described (Lin and Garcia, 2012). C8 solid phase resin (3M, Empore) was used in this case because it performs better with intact proteins compared to the standard C18 resin. Samples were completely dried in a vacuum centrifuge. Subsequently, samples were resuspended in 50uL ammonium bicarbonate, pH 8.0, digested with trypsin as described above for 1 hour (Karch et al., 2016; Lin and Garcia, 2012). Samples were desalted on home-made C18 stage-tip columns and dried to completion in a vacuum centrifuge.

3.4.5: Boronate enrichment of ADP-ribosylated peptides

When indicated, boronate enrichment was performed on 50-100ug of digested histone samples. Samples must be desalted and completely dried before enrichment as water and salt will impede binding to the column. Samples were resuspended in 75uL dimethyl sulfoxide (≥ 99.9% anhydrous, Sigma). Boronate columns (Agilent, Bond Elut PBA, 100mg) were washed with at least 2mL anhydrous dimethyl sulfoxide (DMSO). A P1000 pipettor was used to push the sample through the column (i.e. create a seal at the top of the column with the tip and eject air to push the liquid through). A vacuum manifold could also be used to push volume through the column. Sample was added to the column using a gel loading tip to ensure that the whole sample goes into the column. 50uL of dry DMSO was added after to push the sample into the middle of the boronate resin. Sample was incubated on-column for 2 hours at 35˚C. The column was washed with 1 mL anhydrous DMSO followed by 1mL acetonitrile (99.8% anhydrous, Sigma). Samples were incubated on-column with 500uL 1M hydroxylamine in H2O. Columns were

capped to allow longer interaction of the column with the hydroxylamine. Desalting was subsequently peformed with home-made C18 stage-tip columns.

3.4.6: NanoLC-MS/MS

Fused silica microcapillary tubing (75um i.d.; Polymicro Technologies) were pulled in house using a flame to generate a spray tip. Columns were packed with C18 resin (3um, Dr Maisch GmbH, Germany) using a pressure bomb to load the resin. Samples were resuspended to approximately 1ug/uL in 0.1% formic acid in H_2O. Sample was loaded using a Thermo Easy NanoLC 1000 HPLC. Peptides were separated using reversed-phase chromatography over a 60 minute gradient (Buffer A: 0.1% formic acid in water; buffer B: 0.1% formic acid in acetonitrile): 0 to 28%B in 45 minutes, 28 to 90% B in 5 minutes, 90% B for 10 minutes. For all experiments except the digest optimization experiments, the HPLC was coupled to a hybrid linear ion trap-Orbitrap (Orbitrap Fusion, Thermo Scientific). A full MS scan (350-1500 m/z) was acquired in the Orbitrap (60,000 resolution; AGC: 2.0×10^5). MS/MS spectra were collected for ions with an intensity greater than 5.0×10^3 in the Orbitrap (15,000 resolution; AGC: 5.0×10^4). The MS/MS cycle time was 3 seconds, and ions were chosen based on their intensity. Ions with a charge 2-4 were fragmented with HCD (27% collision energy) and ions with charge 4-7 were fragmented with ETD using calibrated charge dependent ETD parameters. Dynamic exclusion was set to 25 seconds after one selection. All raw files are available on the Chorus database (ID: 1377; https://chorusproject.org/).

3.4.7: MS data analysis

All ETD data was processed using Mascot as part of the Proteome Discovery software suite. Xtract was used to deconvolute and de-isotope spectra, and these spectra were searched

using a database containing all human histone sequences. Spectra were filtered to remove precursor peaks, charge reduced precursors, neutral losses, and FT overtones. The data was searched using semiTrypsin as the enzyme with 3 missed cleavages allowed. For digestion optimization experiments, the enyzmes were set to none for pepsin, or chymotrypsin (2 missed cleavages). The precursor mass tolerance was set to 10ppm, and the fragment mass tolerance was specified to 0.4 Da. For experiments seeking to identify ADP-ribosylation sites, variable modifications were set to protein N-terminal acetylation, methionine oxidation, and hydroxamic acid (on Asp or Glu). For digestion optimization experiments, variable modifications were specified: acetyl (K), dimethyl (K or R), methyl (K or R), protein N-terminal acetylation, and trimethyl (K). The target false discovery rate (FDR) was set to 0.01, and all spectra containing modifications were manually validated.

pFind Studio (version 3) was used to process HCD data (Wang et al., 2007). A database containing all known human histone sequences was uploaded to search raw data files. Trypsin was selected as the digestion enzyme with up to 4 missed cleavages. The precursor tolerance was set to 10ppm and the fragment tolerance was set to 0.04 Da. Variable modifications were included: methionine oxidation, protein N-terminal acetylation, and hydroxamic acid on Asp or Glu. All identified spectra were manually validated.

Label-free quantification was performed by summing the intensities of the peptides containing the modification of interest and dividing it by the total intensity of all peptides containing that modification site in all of its charge states and modified forms. The intensities were determined by measuring the area under the curve of the monoisotopic peak of the peptide.

CHAPTER 4. Development of MD- and TD-HDX-MS/MS methodology and application to histone complexes

4.1: Introduction to HDX-MS/MS

Hydrogen deuterium exchange (HDX) is a useful method used to monitor structure and stability of protein molecules (Englander, 2006) (Figure 4.1). This method can be used to study a variety of sample types, including single proteins, protein aggregates, protein complexes, protein/ligand complexes, and protein/nucleic acid complexes. Recently, some work has been done to expand this method to accommodate small molecules as well, although this work is less common (Schneider et al., 2013).

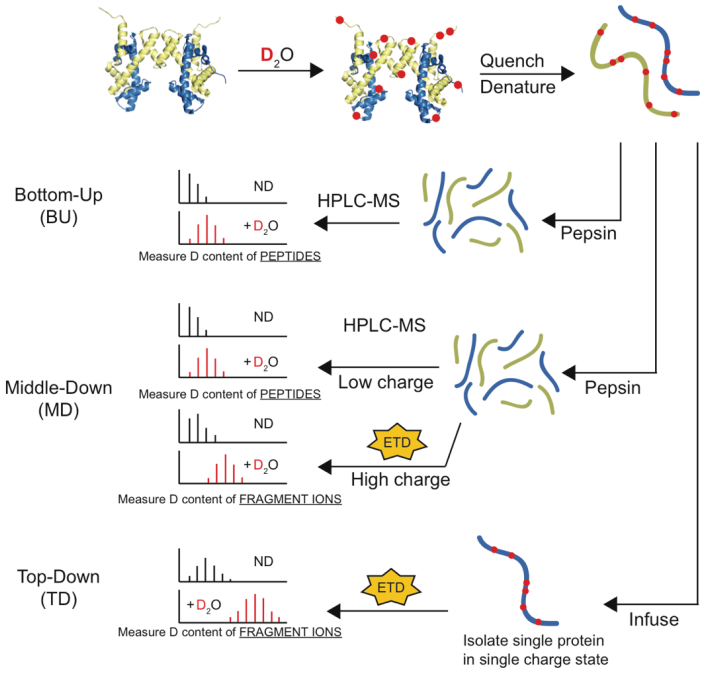


Figure 4.1. Schematic of HDX-MS Experiments. Proteins are diluted into D_2O buffer and allowed to exchange over time. Aliquots are quenched at different timepoints and deuterium content is measured by mass spectrometry. In bottom-up experiments, proteins are digested with

pepsin, separated by RP-HPLC, and deuterium content of the intact peptides are measured. In middle-down experiments, the same sample processing occurs but highly charged peptides can be targeted for fragmentation to obtain more resolved information. In top-down experiments, intact proteins are infused into the mass spectrometer and fragmented by ETD. Deuterium content of fragment ions are measured. ND = non-deuterated.

In an HDX experiment, proteins of interest are diluted into buffer containing heavy water (deuterium oxide, D_2O) over a course of time. Buffers are made such that the proteins are in native conditions and properly folded. Amide protons on the backbone of the protein can then exchange for deuterium atoms from the solvent. However, amide protons may be protected from exchange if they are not readily accessible by solvent or if they are participating in a hydrogen bond, as is the case for secondary structures. However, given that proteins are dynamic molecules, transient unfolding events can break hydrogen bonds and lead to deuteration of the protein backbone. More stable secondary structures will undergo fewer transient unfolding events compared to less stable structures and will therefore have a slower rate of deuterium exchange. The rate of deuterium incorporation therefore acts as a proxy for protein structure and stability, with faster rates indicating less stable structures. HDX can also be used to map binding interfaces with small molecule ligands, other proteins, or nucleic acids as amino acids involved in hydrogen binding at the interface will be more protected from exchange compared to unbound. Deuterium incorporation can be measured by mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR) (Englander, 2006; Krishna et al., 2004).

The traditional HDX-MS experiment involves performing the exchange reaction for sample of interest over time points spanning several orders of magnitude in duration. After a time point is reached, the sample is “quenched” with low pH buffer containing concentrated denaturant. A final pH of ~2.25 is desirable as the exchange reaction is at its minimal rate at this pH (Walters et al., 2012). Then, samples are digested with pepsin and the resulting peptides are desalted and separated by reversed-phase high performance liquid chromatography (RP-HPLC)

before being sprayed into the mass spectrometer. The addition of denaturant in the quench buffer facilitates digestion with pepsin. Generally, the sample processing steps are performed online in a cooled apparatus containing a sample loop for injection, a pepsin column for digestion, a trap column for online desalting, and a C18 column for separation (Englander, 2006).

One important consideration in any type of HDX-MS experiment is back-exchange (Walters et al., 2012). Back-exchange occurs when a deuterated sample comes into contact with water from HPLC and/or desalting buffers. Hydrogen atoms from the water in the buffers can exchange with deuterium from the sample. Back-exchange may not occur at an equal rate across the sequence of the protein, and so it is very critical to include a fully-deuterated control in each experiment. Fully-deuterated controls are made by solvating your protein of interest in D_2O containing a high concentration of denaturant to unfold proteins. Since there is no longer protection from secondary structures, the protein should be fully deuterated. To further ensure a maximum level of deuteration, samples are often exchanged at 37˚C for several days. The deuteration level of this control can then be used to normalize the data from time points to ensure that differences in back-exchange do not affect the interpretation of results.

Back-exchange can be minimized by tightly controlling temperature and pH. Hydrogen/deuterium exchange is minimized at pH 2.25. Ensuring that the addition of quench buffer generates a final pH of 2.25 and that the HPLC and desalting buffers are also precisely this pH will therefore reduce the extent of back-exchange. Back-exchange can also be minimized by ensuring that temperatures are as low as possible (Walters et al., 2012). The exchange reaction increases by a factor of 10 per 22˚C (Englander, 2006). Generally, samples are flash-frozen immediately after addition of quench buffer and thawed in ice water before injection, and a cooling apparatus is used to ensure that digestion, desalting, and separation occur at temperatures between 0-5˚C.

After samples are processed online, MS is often used to measure the deuterium content of the peptides. A non-deuterated sample is run first to identify the peptides and their retention

times, and this information can be used to identify the deuterated peptides in subsequent runs. Often times, several non-deuterated samples are run, generating exclusion lists after each, to get deeper coverage of the sample. This method, where proteins are digested into peptides prior to analysis by MS is called bottom-up mass spectrometry (BU-MS) (Figure 4.1). Thus far, this BU-HDX-MS method has been the primary HDX method of choice by the scientific community.

However, one major drawback of the traditional BU-HDX-MS method is that the resolution of the information is as good as the length of the peptide, meaning that deuterium content cannot be localized within the peptide (Kaltashov et al., 2009). To clarify with an example, let's say there's a peptide containing 6 amino acids that was measured to contain the equivalent of three deuterons. In this method, it is impossible to localize those three deuterons to specific amino acids; for example, it could be that the first three residues are fully deuterated, or that all six residues are 50% deuterated, or that every other amino acid is fully deuterated, etc. If a user is able to obtain many overlapping peptides, there is a possibility of achieving higher resolution by subtracting the masses of the overlapping peptides (Gessner et al., 2017; Kan et al., 2013; Mayne et al., 2011). However, site-specific localization of deuterium content is very rare and difficult to achieve in the traditional BU-HDX-MS method.

Tandem mass spectrometry (MS/MS) offers a potential solution to the resolution problem of BU-HDX-MS. If peptides are fragmented, the masses of the resulting fragment ions could be used to localize deuterium to specific residues. However, one major hurdle to this method is the phenomenon of scrambling. In scrambling, hydrogen and deuterium atoms can migrate along the protein backbone in the gas phase, essentially randomizing signal. Although the exact mechanism of scrambling is not known, researchers have determined that it occurs when peptide ions reach a high vibrational energy. Unfortunately, traditional collision-based fragmentation methods, such as collision induced dissociation (CID) and higher-energy collisional dissociation (HCD), lead to nearly 100% scrambling of peptides and are therefore not amenable to HDX-MS experiments (Jørgensen et al., 2005a, 2005b; Kaltashov et al., 2009). The occurrence of

scrambling using CID and HCD fragmentation has precluded the use of tandem mass spectrometry in HDX-MS for decades.

However, it has recently been shown that electron transfer fragmentation methods including electron transfer dissociation (ETD) and electron capture dissociation (ECD) can retain deuterium labeling in the gas phase (Rand and Jørgensen, 2007; Rand et al., 2009; Zehl et al., 2008). ETD is generally more common due to the fact that it is compatible with a chromatographic timescale while ECD scans have a long acquisition time. Given that ETD and ECD rely on transfer of an electron, high vibrational energies are not required and scrambling can be reduced to negligible levels. However, other instrument parameters, such as the isolation window, can impart vibrational energy and therefore must be optimized to "gentle" conditions where scrambling does not occur. Several groups have succeeded in tuning instrument conditions such that scrambling is negligible (Masson et al., 2017; Rand and Jørgensen, 2007; Rand et al., 2009; Zehl et al., 2008). However, optimal instrument parameters likely vary between instruments, and so the HDX user should optimize the parameters on their own instrument.

The discovery that ETD fragmentation can minimize scrambling allows for different types of HDX-MS methods to be developed. One such method is middle-down HDX-MS/MS (MD-HDX-MS/MS) (Figure 4.1). In this method, peptides that are ETD-amenable can be targeted for fragmentation, and deuterium content of the fragment ions can be measured to allow for more resolved information. However, given that ETD is an electron transfer mechanism, only peptides with sufficient charge will be amenable to fragmentation (usually charge +5 or higher). Therefore, more resolved information can only be obtained for portions of the protein. We termed this method MD-HDX-MS/MS because it enables analysis of long peptides, as these peptides generally carry more charge and are therefore better suited for ETD fragmentation. The set-up for MD-HDX-MS/MS is nearly identical to that for BU-HDX-MS, with the exception that instrument parameters must be optimized to reduce scrambling and that some peptides are fragmented with ETD.

Top-down HDX-MS/MS (TD-HDX-MS/MS) can also be conducted. In this method, digestion and separation steps are omitted, and the entire sample is infused into the mass spectrometer. One charge state of one protein ion can then be targeted for fragmentation, and the deuterium content of the fragment ions can be measured. Generally, one minute of fragmentation provides more than enough spectra to average for to obtain one high-quality averaged spectrum for analysis. TD-HDX-MS/MS offers several advantages over BU-HDX-MS and MD-HDX-MS/MS: (1) full coverage of the protein is guaranteed, whereas digestion may lead to gaps in sequence coverage, (2) up to site-specific resolution can be obtained for the entire protein depending on the efficiency of fragmentation, (3) about 10- to 20-fold less instrument time is required, and (4) back-exchange is more easily minimized due to less contact with water-containing buffers. However, MD-HDX-MS/MS can accommodate more complex samples and has a smaller sample requirement.

Despite the great utility of HDX coupled to ETD MS/MS, very few studies have utilized these platforms to monitor protein structure and function. BU-HDX-MS/MS has only been accomplished in a handful of studies, all of which have focused on a single protein or oligomers of a single protein (Table 4.1) (Abzalimov et al., 2013; Huang et al., 2011; Landgraf et al., 2012; Masson et al., 2017; Pan et al., 2016; Rand et al., 2009). TD-HDX-MS/MS has been slightly larger in scope but is still mainly limited to small (< 200 amino acids) already well-characterized proteins as proof-of-principle experiments (with exceptions). For example, the Borchers group has used TD-HDX-MS/MS to study antibodies (Table 4.1) (Abzalimov et al., 2009; Going et al., 2016; Hoerner et al., 2005; Pan and Borchers, 2014; Pan et al., 2008, 2009, 2010, 2011, 2012, 2014, 2016; Sterling and Williams, 2010; Xiao and Kaltashov, 2005). Together, these initial studies demonstrate the great power of ETD-HDX-MS/MS techniques to monitor site-specific deuterium incorporation in unprecedented detail and also highlight the potential of these methods to be applied to more complex samples.

Table 4.1. Previous HDX-MS/MS studies.

	#	Last Author	Year	Protein	Fragmentation	Protein Size	Type of experiment
MIDDLE-DOWN	1	Jørgensen	2009	β2-microglobulin	ETD	99 residues	Proof-of-principle
	2	Gross	2011	Oligomers of apolipoprotein E	ETD	299 residues (34kDa)	Measure differences in mutant and WT
	3	Griffin	2012	PPARG ligand binding domain	ETD	Not given	Map ligand binding, proof-of-principle
	4	Kaltashov	2013	N-lobe of human serum transferrin	ECD	37kDa	Measure protein conformational dynamics
	5	Williams	2017	PI3K	ETD	Not given	Map inhibitor binding sites
TOP-DOWN	6	Kaltashov	2005	Ubiquitin	CAD	76 residues	Analyze dynamic properties of a partially folded conformation
	7	Kaltashov	2005	CRABP I	CAD	18kDa	Measure conformational stability
	8	Konermann	2008	Ubiquitin	ECD	76 residues	Proof-of-principle
	9	Konermann	2009	Myoglobin (Mb)	ECD	17kDa	Measure with and without heme, proof-of-principle
	10	Kaltashov	2009	CRABPI	ETD	18kDa	Proof-of-principle
	11	Borchers	2010	Apomyoglobin (aMb)	ECD	153 residues	Measure folding intermediates
	12	Williams	2010	ubquitin	ETD	76 residues	Effects of supercharging; proof-of-principle
	13	Konermann	2012	Aβ oligomers	ECD	40 residues	Study protein conformers
	14	Kaltashov	2013	Ubiquitin (Ub)	ECD	72 residues	Analyze a protein conformer
	15	Kaltashov	2014	β2-microglobulin	ECD	99 residues	Test disulfide reduction in HDX
	16	Borchers	2014	Calmodulin; Herceptin	ECD	17 and 150 kDa	Proof-of-principle
	17	Borchers	2014	Interferon α2a and variants	ECD	~60kDa	Compare structure of variants
	18	Borchers	2015	Herceptin (HER)	ETD	150kDa	Measure effects of ligand binding
	19	Williams	2016	RNase A, ubiquitin, myoglobin	ETD	14, 8.5, and 16.7kDa	Effects of supercharging on scrambling
	20	Borchers	2016	bevacizumab (BEV)	ETD	Heavy chain: 50kDa	Measure differences between biosimilars

References: 1: (Rand et al., 2009); 2: (Huang et al., 2011); 3: (Landgraf et al., 2012); 4: (Abzalimov et al., 2013); 5: (Masson et al., 2017); 6: (Hoerner et al., 2005); 7: (Xiao and Kaltashov, 2005); 8: (Pan et al., 2008); 9: (Pan et al., 2009); 10: (Abzalimov et al., 2009); 11: (Pan et al., 2010); 12: (Sterling and Williams, 2010); 13: (Pan et al., 2012); 14: (Wang et al., 2013); 15: (Wang and Kaltashov, 2014); 16: (Pan and Borchers, 2014); 17: (Pan et al., 2014); 18: (Pan et al., 2015); 19: (Going et al., 2016); 20: (Pan et al., 2016).

The goal of this portion of my thesis is to expand upon the existing ETD-based HDX-MS/MS technology to accommodate protein complexes and protein/DNA complexes for the first time. I chose to study nucleosomes and sub-nucleosomal histone complexes to achieve this goal because they can exist in a variety of complexes, are amenable to ETD fragmentation, are difficult to study by traditional BU-HDX-MS, and have significant biological importance.

As discussed in Chapter 1, histone proteins are critical for maintaining nuclear integrity. The N-terminal tail domains are of particular interest in the field because they are the site of the majority of post-translational modification, are directly involved in forming higher order chromatin structure, and serve as the binding site for many enzymes and proteins involved in chromatin regulation (Jenuwein and Allis, 2001; Portela and Esteller, 2010; Zhao and Garcia, 2015). Despite the great interest in histone tails, their structure and dynamic properties are currently unknown, mainly due to the lack of a rigorous method to address this question.

However, there are several illuminating studies that have aimed to probe histone tail structure and dynamics that provide a great starting point for this analysis. For example, the Akashi group performed two studies to monitor tail the gas-phase structure of the nucleosome core particle (histone octamer without DNA), $(H3/H4)_2$ tetramers, and H2A/H2B dimers using ion mobility mass spectrometry and molecular dynamics (MD) simulations (Saikusa et al., 2013, 2015). The results indicated that histone tails are likely highly unstructured and can occupy many different conformations. However, many other MD studies have been conducted on histones in various contexts, and there has not been a lot of consensus about the structure of the histones tails. For example, in contrast to the studies by the Akashi group, the Cui group performed MD simulations on the H3 tail peptide and found that it readily forms an α–helix in two regions of the peptide (Zheng and Cui, 2015). It is likely that the histone construct used and the exact experimental parameters have a large effect on the results. Without the ability to validate these results experimentally, it is difficult to determine which structures are biologically relevant.

Circular dichroism (CD) experiments have revealed that the histone tails may contain some secondary structure in solution. Banères and colleagues analyzed secondary structure content of nucleosomes with and without the tail domains present and found that 60% of H3 and H4 tail residues contain α–helix properties, although the individual contribution of each tail was not determined. They found that the H3 and H4 tails contributed approximately 35% to the total α–helical content of the nucleosome core particle. The group also found that H2A and H2B tails exist as random coils (Banères et al., 1997). However, Wang and colleagues conducted a similar experiment and determined that the tails contribute 17% of the total α–helical content of the octamer, about half of that found in the previous study (Wang et al., 2000). Furthermore, another study analyzed histone H4 tail peptides in isolation with CD and NMR and found no significant secondary structure (Bang et al., 2001). The conflicting results of these studies illustrate the need for a robust and reproducible method to study histone tail structure and stability in solution in greater resolution.

Nucleosome and sub-nucleosomal histone complexes have been previously analyzed by traditional BU-HDX-MS methods without the use of MS/MS to provide a more global analysis of histone structure and stability. One such study by Black et al. analyzed deuterium exchange profiles of $(H3/H4)_2$ and $(CENP\text{-}A/H4)_2$ heterotetramers (Black et al., 2004). CENP-A is an H3 variant located at the centromere. In this study, minor protection was observed in the α-1, α-2, and α-3 helices of canonical H3 and the α-1 and α-2 helices of H4 (Figure 4.2). Notably, the α-N helix of H3 was fully exchanged by 10 seconds. The tail domains of H3 and H4 were detected; however high-resolution information could not be obtained because the identified peptides spanning this region were quite long, ranging from approximately 25 to 50 amino acids. Nonetheless, this study revealed that the tail domains were completely exchanged by 10 seconds, indicating that the tails are likely unstructured and highly dynamic in this context. Additionally, D'Arcy et al. was able to demonstrate that the H2A/H2B dimer is largely unstable, undergoing nearly full exchange within 10 seconds under low salt conditions at room temperature (D'Arcy et al., 2013).

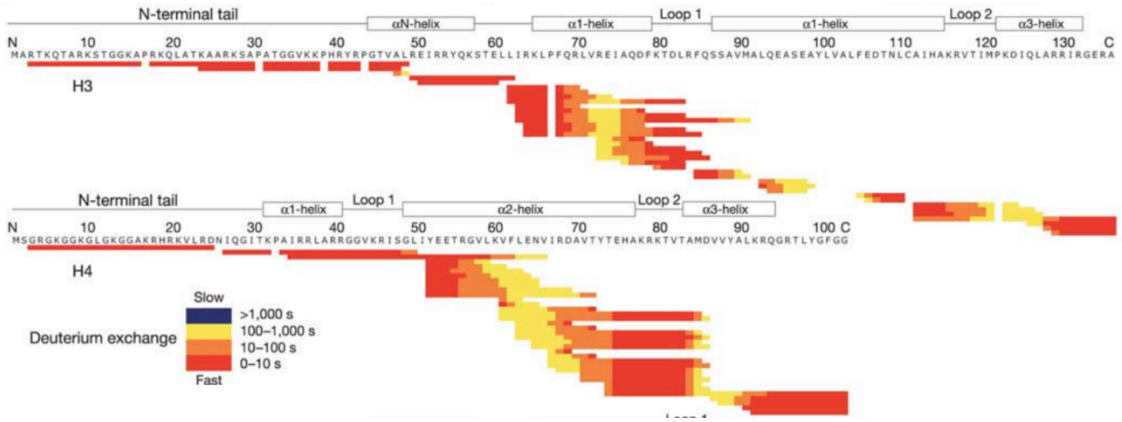


Figure 4.2. HDX Exchange profiles of H3 and H4 in the $(H3/H4)_2$ heterotetramer. $(H3/H4)_2$ heterotetramers were exchanged in D_2O for the indicated time points at room temperature. Each bar represents an identified peptide used in the analysis, and the color indicates at which time point exchange occurs according to the legend. The secondary structural features are displayed above the primary sequence. Figure taken from (Black et al., 2004).

A second study published by Black et al. three years later explored the stability of canonical nucleosomes compared to CENP-A containing nucleosomes (Black et al., 2007). The results of this study demonstrated that H3 and H4 experience a dramatic increase in protection from exchange (>3 order of magnitude) across nearly the entire protein in nucleosomes compared to tetramers. Unfortunately, the tail domains of H4, H2A, and H2B were not detected and therefore could not be studied (Figure 4.3). These tails may not have been detected for several reasons: (1) pepsin may have over-digested the tail regions, yielding peptides that do not retain well on C18 analytical columns or trap columns (tails contain many hydrophilic residues), (2) intact tail peptides may be too long or too highly charged to be identified with CID fragmentation used in this study, or (3) tail peptides may be too low in abundance compared to

other co-eluting peptides to be selected for identification by MS/MS in non-deuterated runs. However, in this study, four H3 peptides were identified that spanned the length of the histone tail and a portion of the α-N helix. These peptides demonstrated significant protection, with deuteration levels reaching only approximately 50% at 10^6 seconds at 37˚C. Unfortunately, given the length of these peptides, the observed protection could not be localized and so it remains unclear whether the H3 tails are protected from exchange in nucleosome context (observed protection could be from the α-N helix), and if so, where this protection occurs. However, the results of these studies have provided a wealth of information about the stability of the histone core in tetramer and nucleosome contexts. These results can also be compared to those obtained in the HDX methods with ETD developed in this thesis to determine if these new methods are adequate for probing histone structure and dynamics.

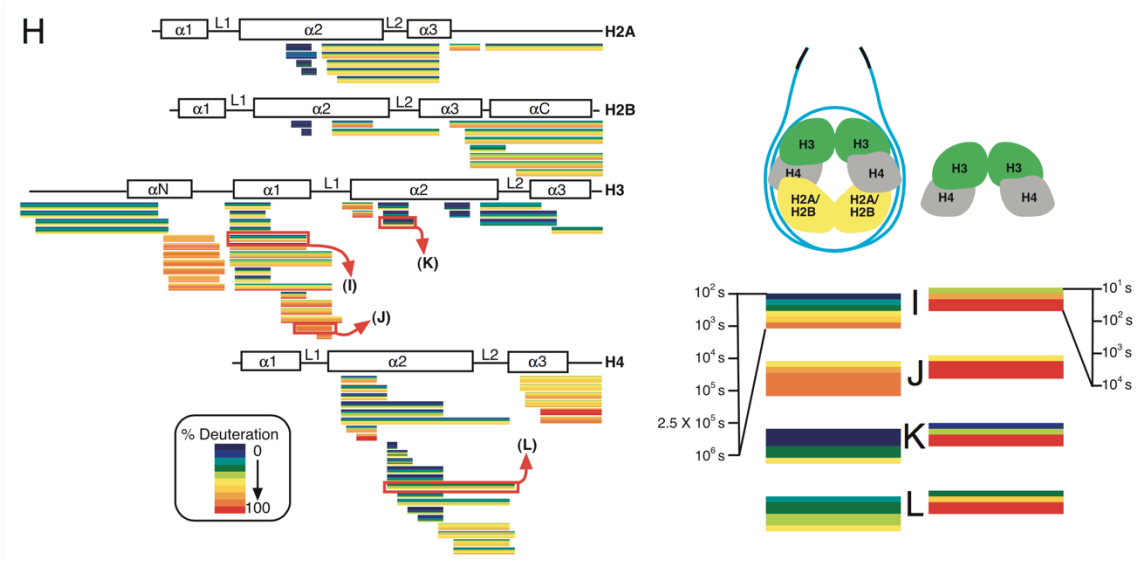


Figure 4.3. Exchange profiles of canonical histone proteins within the nucleosome. Nucleosomes were exchanged in D_2O buffer at 37˚C for the indicated timepoints. Each bar represents a peptide identified and used in the analysis, and each line within the bar represents

the % deuteration of that peptide throughout the time points according to the legend. Tetramer data is from (Black et al., 2004). Figure is taken from (Black et al., 2007).

These HDX-MS results are in line with a study by the Norman group, where pulsed electron-electron double resonance (EPR) spectroscopy coupled to site-directed spin labeling was used to probe histone structure (Bowman et al., 2010). These authors also found that the H3 α-N helix displays increased structural heterogeneity in the tetramer compared to octamer or nucleosome. Furthermore, they also observed that the tetramer exhibits increased structural heterogeneity overall compared to H3 and H4 in nucleosomal context. However, the tail domains were not included in the study.

The goal of this portion of my thesis is to expand upon the existing HDX-MS/MS technology to accommodate protein complexes and protein/DNA complexes. Specifically, I used histone complexes and nucleosomes to demonstrate the power of HDX-MS/MS to monitor protein structure and stability in unprecedented detail. In doing so, I was able to rigorously monitor histone tail dynamics in solution at up to residue-level resolution for the first time and expand the versatility of existing HDX-MS/MS methodology.

4.2: Results

4.2.1: TD-HDX-MS/MS enables robust and reproducible localization of deuterium content in a histone protein complex at near site-specific resolution

As discussed in Chapter 4.1, histone proteins are critical for maintaining nuclear integrity through regulation of many processes including transcription, cellular division, and formation/maintenance of chromatin structure. Guided by chaperone proteins, histones are deposited onto DNA to form nucleosomes in a specific order, with an $(H3/H4)_2$ heterotetramer

being deposited first followed by two H2A/H2B dimers. The crystal structure of the histone tetramer has not yet been solved; however, data from a study using pulsed EPR spectroscopy and site-directed spin labeling indicates that H4 likely contains the three alpha helices present in the nucleosome crystal structure, although H3 may have increased conformational flexibility relative to the nucleosome structure (Bowman et al., 2010). Indeed, previous work by Black et al. demonstrated that histone tetramers experience increased structural rigidity (>3 orders of magnitude) upon incorporation into nucleosomes and exhibit protection in secondary structures (Black et al., 2004, 2007) (Figure 4.3). None of these studies, however, were able to monitor histone tail dynamics in these structures.

Given the critical role of histone tetramers in the nucleus, we sought to develop top-down HDX-MS/MS (TD-HDX-MS/MS) methodology to study histone H4 within the tetramer as well as its monomeric form, representing the first TD-HDX-MS/MS study analyzing a protein complex. Histone tetramers and monomers make a great case study to develop TD-HDX-MS/MS given that they are relatively complex samples, are amenable to ETD fragmentation, are difficult to study by traditional BU-HDX-MS, and have significant biological importance.

Protection from exchange with deuterium occurs when solvent does not have access to amide protons or when amide protons are occupied in hydrogen bonds, as in secondary structures for example. Therefore, slower rates of deuterium exchange can occur upon increased stability (or reduced flexibility) of protein regions including secondary structures, intra- or intermolecular contacts, compaction of the protein, or any combination of these. Therefore, we hypothesized that H4 within the tetramer would be protected from exchange in the secondary structures, while H4 monomers, which are likely unstructured in the absence of their binding partner, would experience faster rates of exchange. However, the structure of H4 monomers have never been explored, and so some secondary structure may exist.

In TD-HDX-MS/MS, samples are desalted and subsequently infused into the mass spectrometer. We desalted samples off-line at 4˚C using home-made C8 stage tip columns. After

desalting, samples were eluted in a buffer containing 75% acetonitrile and 25% (dH_2O, pH 2.25) to promote stable spray. Samples were immediately flash frozen. Samples were thawed and immediately infused into the mass spectrometer using an Advion Triversa Nanomate. This instrument contains an automated sample pick-up mechanism that picks sample up in a plastic tip and brings it to a silicon microfluidics chip that contains 400 nano-electrospray nozzles. The instrument can then push the sample through the tip into a nozzle for stable spray into the mass spectrometer. The design allows for very small sample volumes (<10uL) to be stably infused for up to 30 minutes.

There are two major technical hurdles that must be addressed prior to conducting a TD-HDX-MS/MS experiment: scrambling, whereby proton and deuterium atoms on the backbone can migrate, effectively randomizing signal, and back-exchange, where deuterium atoms on the protein can exchange for protons from desalting and infusion buffers.

As described in Chapter 4.1, scrambling occurs when analyte ions reach high vibrational energy in the gas phase. Previous studies have indicated that using “gentle” MS conditions and fragmentation methods (namely, ETD) can reduce scrambling to negligible levels (Masson et al., 2017; Rand and Jørgensen, 2007; Rand et al., 2009; Zehl et al., 2008). The Jørgensen group developed a scramble “probe” peptide, called P1 (sequence: HHHHHHIIKIIK), that can be used to monitor scrambling (Rand and Jørgensen, 2007). The histidine amino acids at the N-terminus of the peptide have a much faster intrinsic rate of deuterium exchange at the backbone amide, two orders of magnitude greater than that of the isoleucine and lysine residues at the C-terminus. Therefore, the lyophilized peptide can be dissolved in D_2O and allowed to fully exchange, then diluted into H_2O, to allow for selective labeling of the peptide. The amide protons of the histidine residues will exchange for protons from the water while the isoleucine and lysine residues will retain a deuteron at their amide positions. Then, the peptide can be infused into the mass spectrometer under various instrument parameters to determine which settings allow for preservation of the selective labeling and therefore do not promote scrambling.

We monitored scrambling with our TD-HDX-MS/MS instrument set-up under various MS parameters using the same probe peptide 'P1' and were able to minimize scrambling to very low levels (8.2%, Figure 4.4), using the following parameters (isolation window: 10 m/z, S lens RF: 60%, ETD reaction time: 70ms, resolution: 60,000, capillary temperature: 150˚C). Using the harshest conditions (Isolation window: 2 m/z, S lens RF: 70%, ETD reaction time: 150ms, resolution: 60,000, capillary temperature: 150˚C), we obtained 63% scrambling, indicating that our set-up provides very low levels of scrambling under normal operating parameters.

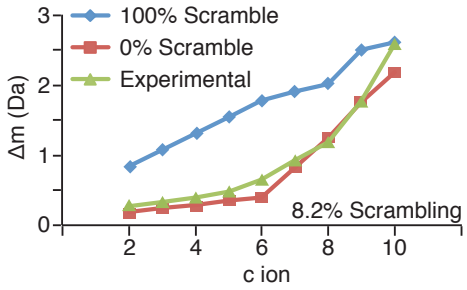


Figure 4.4. Scrambling can be minimized using the TD-HDX-MS/MS setup. The scrambling probe peptide 'P1' was used to monitor scrambling. The peptide was fragmented with ETD and the deuterium content of the resulting c ions was calculated. The non-deuterated centroid was subtracted from the deuterated centroid value to obtain the difference in mass (Δm) and plotted above (green). The theoretical 100% scrambling (blue) and theoretical 0% scrambling (red) Δm values were calculated and plotted.

Back-exchange must also be minimized in the TD-HDX-MS/MS experiment. During the infusion process the sample in the tip can reach very high temperatures given that it is located next to the capillary that is held at 150˚C. Therefore, we constructed a cooling apparatus as previously described by the Jørgensen group (Amon et al., 2012). In this apparatus, copper

tubing is coiled inside an insulated box, which is filled with dry ice (Figure 4.5A). The tubing is attached to a nitrogen tank so that when the nitrogen flows through the box, it cools down to sub-zero temperatures. The other end of the tubing is directed at the tip containing the sample perpendicular to the spray direction to prevent interference with spray stability (Figure 4.5B). To test this cooling apparatus, we infused fully-deuterated histone H2B and measured back-exchange as a shift in the centroid over a 15 minute infusion window. We found that back-exchange was greatly reduced when using the cooling apparatus (Figure 4.5C).

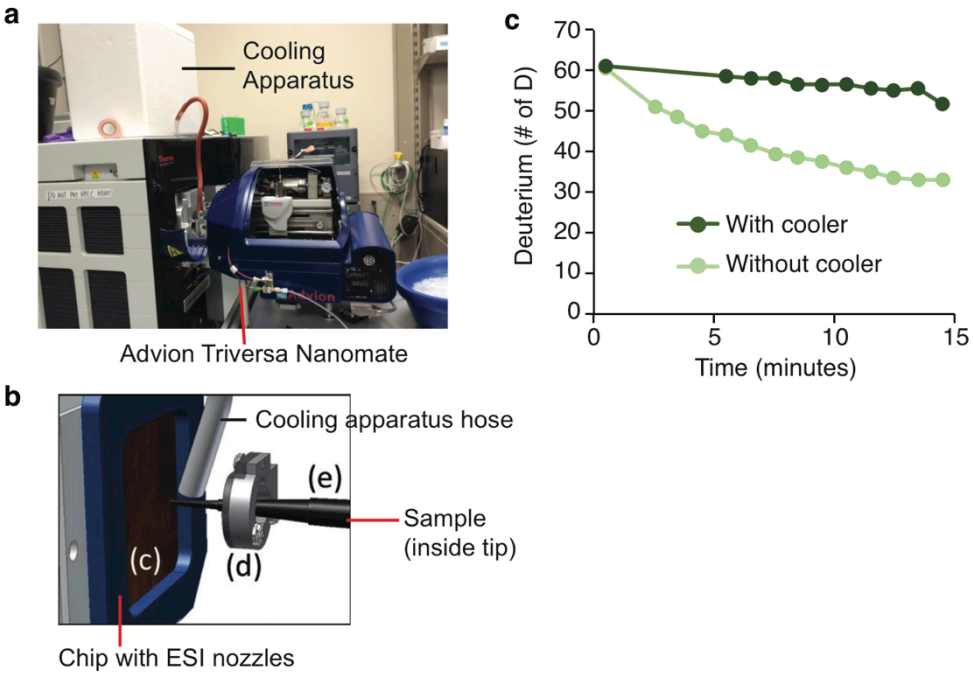


Figure 4.5. TD-HDX-MS/MS set-up with cooling apparatus reduces back-exchange. (A) The Advion Triversa Nanomate is used to deliver sample to the mass spectrometer. A cooling apparatus was constructed to reduce the temperature of the sample that is housed in a spray tip near the ESI chip (B). (C) Fully-deuterated H2B sample was infused into the Thermo Orbitrap

Fusion with and without the cooling apparatus. The deuterium content of the protein was monitored over a 15 minute infusion window. Panel B is modified from (Amon et al., 2012).

After ensuring that back-exchange and scrambling levels were minimized, we made samples to analyze by TD-HDX-MS/MS. To this end, histones corresponding to the human sequences of H3.1 and H4 were expressed and purified in E. coli. Histone tetramers were reconstituted by salt dialysis. H4 monomers and tetramers were incubated in D_2O for varying time points, including 10^1s, 10^2s, 10^3s, and 10^4s at 4˚C, to allow for deuterium exchange for amide protons on the protein backbone. Fully deuterated samples were also made in which samples were exchanged in the D_2O buffer containing a large amount of deuterated denaturant (2.5M guanidine-d_5 DCl) for several days at 37˚C to ensure near maximal deuteration levels. Experiments were performed in triplicate. H4 in charge state +15 was targeted for ETD fragmentation, and deuterium content of the fragment ions was calculated using HDExaminer software (Sierra Analytics, version 2.5) and ExMS2 software developed by the Englander lab (Figure 4.6).

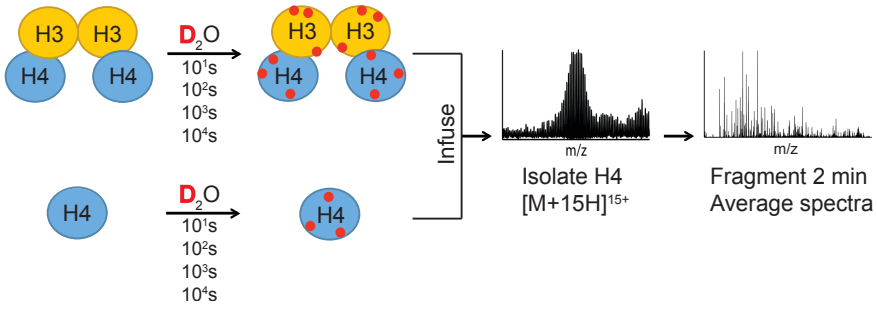


Figure 4.6. Experimental scheme for TD-HDX-MS/MS experiments. Histone H4 monomers and $(H3/H4)_2$ tetramers were exchanged in D_2O buffer for the indicated time points. Samples were quenched and flash frozen. Samples were later thawed and infused into the Thermo

Orbitrap Fusion, where H4 in the +15 charge state was isolated and fragmented for 2 minutes. Spectra were averaged to obtain one high quality MS/MS spectrum for analysis.

We first analyzed the reproducibility of deuterium content measurements to ensure that TD-HDX-MS/MS is capable of robust deuterium measurements. High reproducibility can lend higher confidence to the observed results and enable the user to determine statistically significant differences in deuterium content more easily. To this end, we analyzed the coefficient of variation (CV) of the deuterium content of each detected fragment ion between the three experimental replicates (Figure 4.7). The results show that the CV values are very low, with the median CV value for the monomer and tetramer experiments being below 10% and 5%, respectively. Generally speaking, proteins with a high degree of disorder or flexibility have have a higher degree of variance in deuterium content measurements in HDX experiments because they occupy a larger number of different structures, thereby widening the isotope distribution in HDX experiments and lowering the signal. This makes it more difficult to accurately measure the deuterium content of fragment ions or peptides from disordered proteins. Indeed, we see that the H4 monomer, which we expect to be less structurally rigid than H4 in tetramer context has a slightly higher median CV value. These results indicate that the sample processing steps and the TD-HDX-MS/MS platform enables highly reproducible measurements of deuterium incorporation.

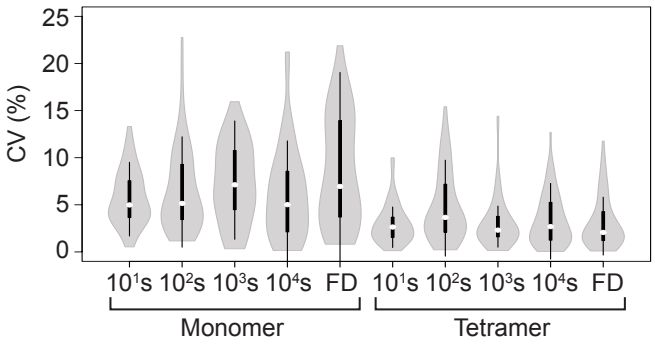
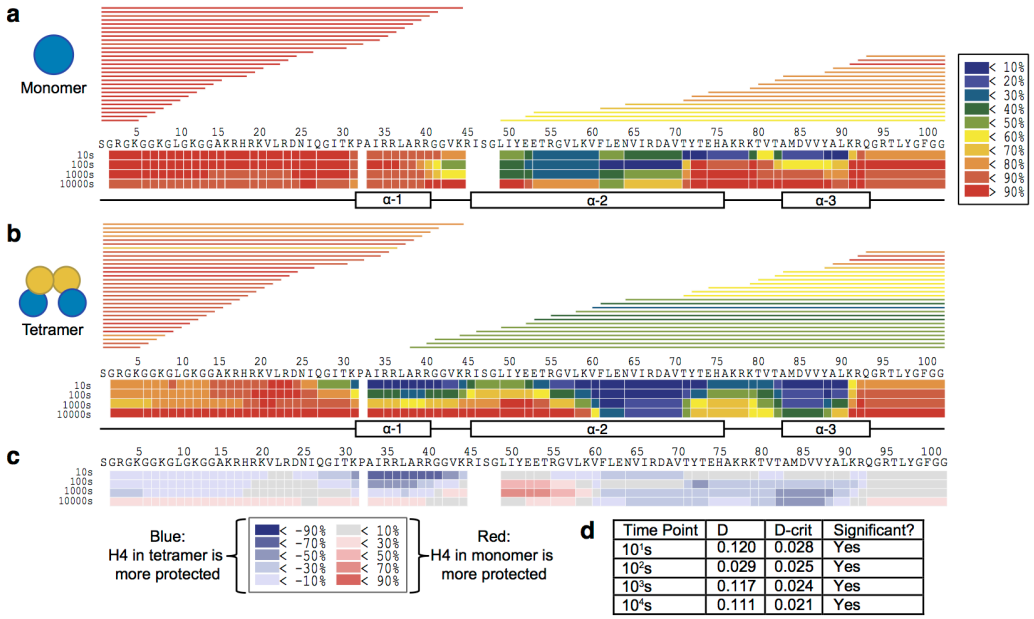


Figure 4.7. Reproducibility of TD-HDX-MS/MS results. The CV values of the deuterium content of each fragment ion were calculated for the three experimental replicates and plotted. White circles show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; polygons represent density estimates of data and extend to extreme values. Plot was generated using BoxPlotR (Spitzer et al., 2014).

Figure 4.8 displays the results of the TD-HDX-MS/MS experiment. We were able to achieve near site-specific resolution for H4 in the monomer and tetramer constructs (28 or 36 site specific sites, respectively). H4 was found to have strikingly different exchange profiles in the monomer compared to tetramer as displayed in Figure 4.8C. Notably, the H4 monomer was found to have protection in some regions, particularly in the α-2 helix; however, in most regions, H4 within the tetramer was found have a larger degree of protection than the monomer as expected. Furthermore, the H4 tail in the tetramer was found to have a small degree of protection from exchange at the N-terminus.



Time Point	D	D-crit	Significant?
10^1s	0.120	0.028	Yes
10^2s	0.029	0.025	Yes
10^3s	0.117	0.024	Yes
10^4s	0.111	0.021	Yes

Figure 4.8. TD-HDX-MS/MS can measure deuterium levels at near site-specific resolution. Heat maps display the average % deuteration of each residue or group of residues according to the legend for H4 in (A) monomeric form or (B) within the tetramer from three experimental replicates. Each bar above the heat map indicates a fragment ion that was used for the analysis, and color of the bar indicates the % deuteration according to the legend. The secondary structure of H4 within the nucleosome is indicated below the heat map. (C) Heat map demonstrating the difference in exchange between monomer and tetramer. Deuterium % of the tetramer was subtracted from the monomer, and the results are shown according to the legend. Blue indicates that the tetramer is more protected and red indicates that the monomer is more protected from exchange. Proline residues are shown as a gap because they do not contain a backbone amide. (D) K-S two-sample test results to determine if H4 has different exchange profiles in the monomer compared to tetramer at each time point. The maximum distance between the cumulative distribution of deuterium content (%) per residue of H4 in monomer and tetramer contexts was calculated (D) and compared to the calculated D-critical value (D-crit).

To determine whether the observed differences in exchange profiles of H4 in monomer versus tetramer context are statistically different, we performed a Kolmogorov-Smirnov (K-S) two-sample t-test. The K-S test is a non-parametric test that compares the cumulative distributions of two different data sets. To determine if the data sets are different, the K-S test calculates the largest difference between the two cumulative distributions of the data sets (D) and compares this value to the largest difference tolerated under the null hypothesis (D-critical; D-crit), which states that cumulative distributions are the same. Therefore if D is greater than D-critical, the null hypothesis is rejected, and the difference between the data sets is considered statistically significant. After performing the K-S test on our data, we found that the differences between the exchange profiles of H4 in monomer compared to tetramer contexts are statistically significant (10s: D = 0.120, D-crit = 0.028, 100s: D = 0.029, D-crit = 0.025, 1,000s: D = 0.117, D-crit = 0.024, 10,000s: D = 0.111, D-crit = 0.021).

The regions of protection found in H4 within the tetramer map to predicted secondary structures from the nucleosome crystal structure (Figure 4.9). The structure shown contains *Xenopus laevis* histone sequences; however, the H4 sequence is identical to human and the H3 sequence is highly similarity (99% identical). It is important to note that the displayed tail domains were not solved in the crystal structure (electron density was too low), but rather modeled into the structure afterwards. These results demonstrate that TD-HDX-MS/MS can easily accommodate a histone complex and provide highly resolved exchange profiles. We are able to get site-resolved information for the H4 tail domain, which has evaded rigorous analysis for decades.

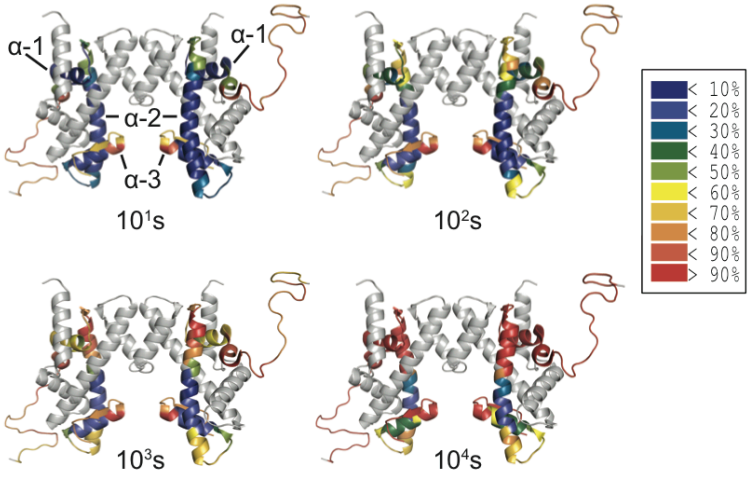


Figure 4.9. TD-HDX-MS/MS reveals that regions of protection for H4 within the tetramer map to predicted secondary structures from the nucleosome crystal structure. Exchange profiles for H4 from the tetramer context are mapped to the crystal structure of the nucleosome (with DNA, H2A, and H2B hidden) (PDB: 1kx5). The color of the residues indicates the % deuteration for that residue as indicated by the legend. Histone H3 is colored in grey.

One inherent drawback of TD-HDX-MS/MS as conducted here is that omission of the RP-HPLC separation step limits the sample complexity that can be accommodated. Figure 4.10 shows the full MS spectrum of the tetramer in non-deuterated and fully-deuterated forms. The isotopic envelopes of H3 and H4 in different charge states overlap significantly in the fully-deuterated spectrum, making it very challenging to select a single species for fragmentation. If multiple species are selected together, the MS/MS spectrum will contain a lot of overlap from the many different fragment ions present. Therefore, it will be very challenging to adapt this methodology to accommodate a more complex sample without the use of chromatography.

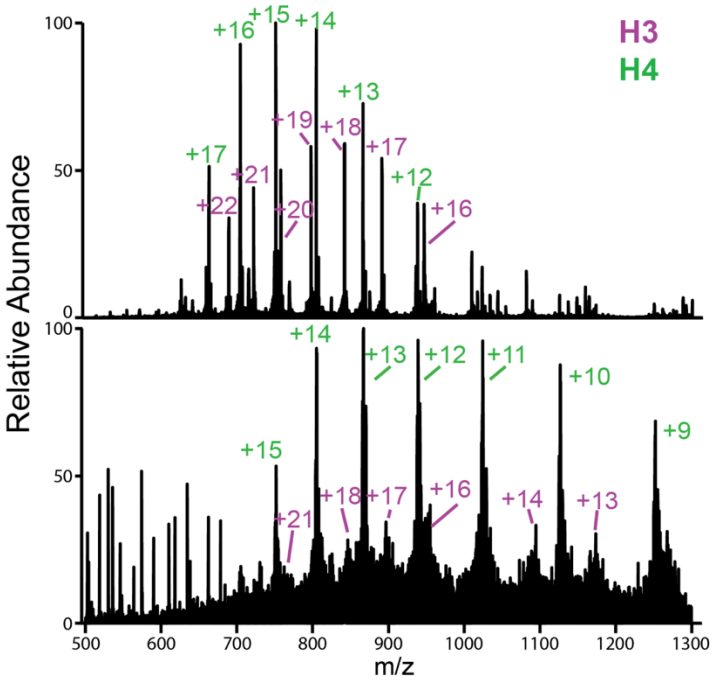


Figure 4.10. Full MS spectra of H3/H4 tetramer demonstrates high degree of overlap upon full deuteration. The full MS spectra are shown for non-deuterated (top) and fully deuterated (bottom) H3/H4 tetramer collected during TD-HDX-MS/MS experiments. Peaks containing H3 are labeled in purple; H4 is labeled in green.

4.2.2: MD-HDX-MS/MS enables high-resolution analysis of histone tail domains in tetramer and nucleosome context

Bottom-up HDX-MS (BU-HDX-MS) and middle-down HDX-MS/MS (MD-HDX-MS/MS) employ digestion with trypsin and separation by RP-HPLC prior to analysis. As such, these platforms can handle more complex samples than TD-HDX-MS/MS, which does not typically employ separation techniques. The platform for BU-HDX-MS and MD-HDX-MS/MS are similar. In both, after the exchange reaction is completed, samples are injected into a cooled online sample processing system containing a pepsin column for digestion, a trap column for desalting, and a C18 analytical column for separation. In BU-HDX-MS, all peptides are analyzed at the MS1 level,

meaning that they are not fragmented to get more site-resolved information. In MD-HDX-MS/MS, ETD-amenable peptides can be fragmented to get more site-resolved information while peptides that cannot be fragmented by ETD are analyzed at the intact level.

Given that these platforms can handle more complex samples compared to TD-HDX-MS/MS, we sought to optimize the MD-HDX-MS/MS approach to accommodate $(H3/H4)_2$ tetramers and nucleosomes (Figure 4.11). This study represents the first heterogeneous protein complex and protein-DNA complex to be analyzed by MD-HDX-MS/MS. For comparison purposes, we analyzed these data sets without the ETD data (analogous to BU-HDX-MS) and with the ETD data (MD-HDX-MS/MS).

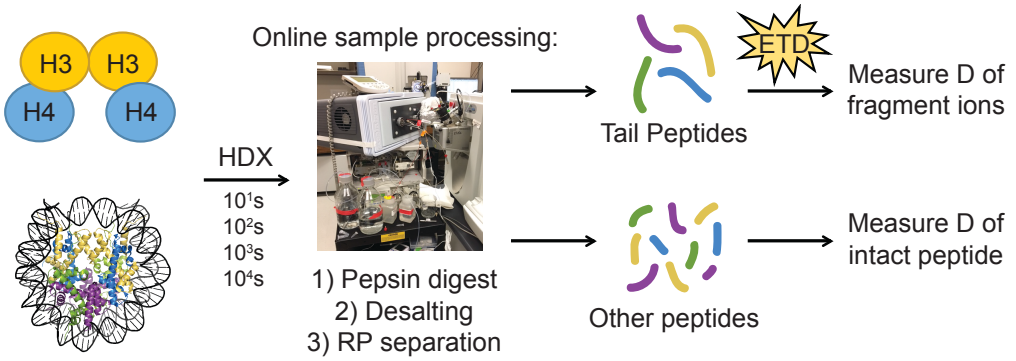


Figure 4.11. Experimental scheme for MD-HDX-MS/MS experiments. $(H3/H4)_2$ tetramers and nucleosomes were exchanged for the indicated time points in D_2O. Samples were injected into a cooled online sample processing system containing a pepsin column for digestion, a trap column for desalting, and C18 analytical column for separation. Tail peptides were targeted for fragmentation by ETD and all other peptides were analyzed at the intact peptide level.

To this end, histones corresponding to the human sequences of H3.1, H4, and canonical H2A and H2B were expressed and purified in E. coli. A His-tag was engineered at the N-terminus

of histone H2A to enable purification with a nickel column. A majority of the His-tag was cleaved off after purification, but a small 4-amino acid tag remains at the N-terminus. Histone tetramers or nucleosomes were reconstituted by salt dialysis. Nucleosomes were made with 197bp DNA sequence corresponding to the well-known 601 nucleosome positioning sequence. Tetramers and nucleosomes were incubated in D_2O for varying time points, including 10^1s, 10^2s, 10^3s, and 10^4s at 4˚C, to allow for deuterium exchange of amide protons on the protein backbone. Experiments were performed in at least triplicate. Each tail was identified in the non-deuterated control and was targeted for ETD fragmentation in deuterated sample runs. In most cases a deuterated sample had to first be run to determine which masses to target. Results were analyzed using HDExaminer software (Sierra Analytics, version 2.5).

Given that the MS set-up for MD-HDX-MS/MS is different than that for TD-HDX-MS/MS, we first sought to ensure that scrambling was not occurring on our platform. To this end, we analyzed the H2A tail peptide, which has the highest charge density of the peptides analyzed and is therefore the most likely to scramble. We calculated the theoretical deuterium content of each fragment ion under 100% scrambling conditions based on how many labile hydrogen atoms are present as has been done by the Williams group (Masson et al., 2017). We plotted these values along with the experimentally obtained deuterium content values (Figure 4.12). To determine whether the experimental and theoretical values are different, we performed a Kolmogorov-Smirnov (K-S) two-sample t-test. The K-S test is a non-parametric test that compares the cumulative distributions of two different data sets. To determine if the data sets are different, the K-S test calculates the largest difference between the two cumulative distributions of the data sets (D) and compares this value to the largest difference tolerated under the null hypothesis (D-critical; D-crit), which states that cumulative distributions are the same. Therefore if D is greater than D-critical, the null hypothesis is rejected, and the difference between the data sets is considered statistically significant. After performing the K-S test on our data, we found that the difference between the theoretical and experimental deuterium content values are statistically

significant (D = 0.291, D-crit = 0.144), indicating that scrambling is not occurring in our MD-HDX-MS/MS platform.

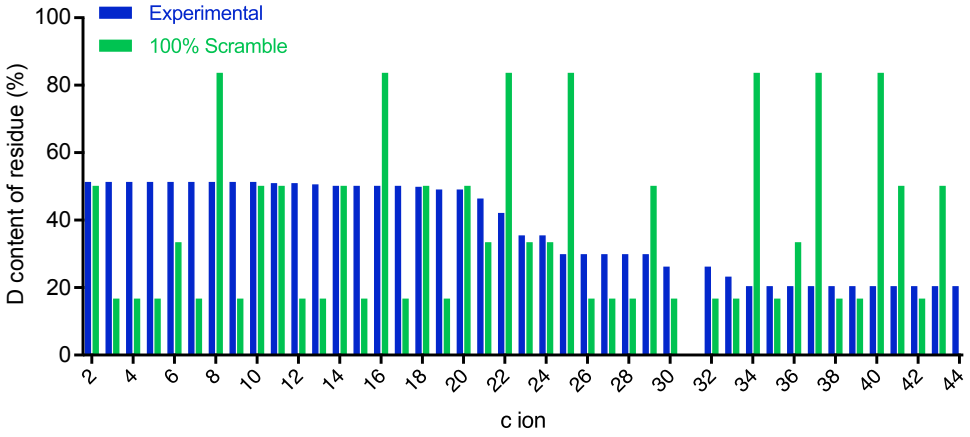


Figure 4.12. Scrambling does not occur on the MD-HDX-MS/MS platform. The H2A tail peptide was targeted for fragmentation and the deuterium content of each fragment ion was calculated using HDExaminer (version 2.5). The theoretical deuterium content values under 100% scrambling was calculated for each fragment ion and compared to experimentally obtained values to determine if scrambling is occurring. A K-S test determined that the difference between the theoretical and experimental values are statistically significant (D = 0.291; D-crit = 0.144).

After determining that scrambling is not occurring, we next analyzed the data without considering the ETD data (analogous to BU-HDX-MS). These analyses will provide a benchmark to assess the improvement in data quality afforded by the use of ETD fragmentation. We first assessed the reproducibility of deuterium content measurements between experimental replicates for each peptide to ensure that this platform enables robust deuterium content measurements. To this end, we calculated the CV of deuterium content measurements between

the experimental replicates for each identified peptide. We found that the deuterium content measurements are highly reproducible, with the median CV value being below 10% for each protein (Figure 4.13).

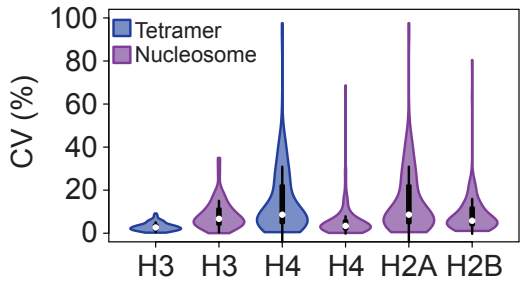
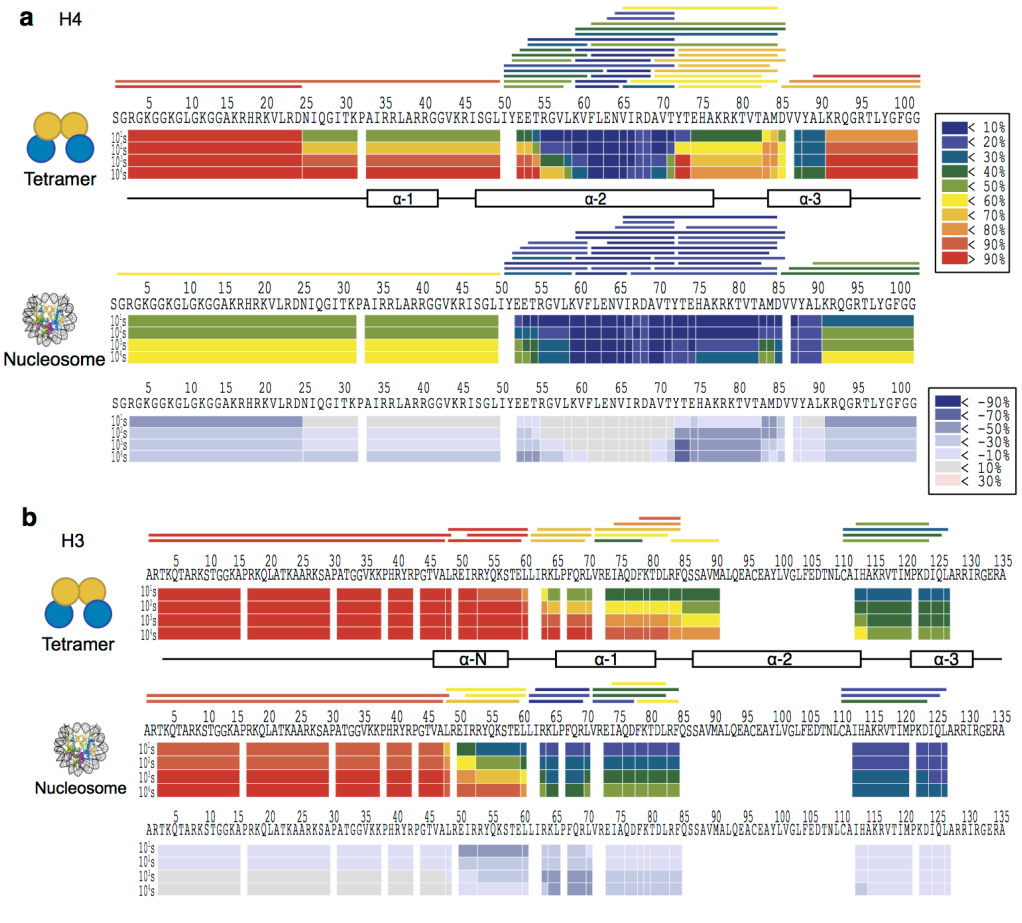


Figure 4.13. Deuterium content measurements in the BU-HDX-MS platform are highly reproducible. The CV values of the deuterium content of each peptide were calculated for the three experimental replicates and plotted. White circles show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; polygons represent density estimates of data and extend to extreme values. Plot was generated using BoxPlotR (Spitzer et al., 2014).

Figure 4.14 shows the results of the BU-HDX-MS analysis for histone H3 and H4 in the tetramer and nucleosomal contexts. Both H3 and H4 experience a dramatic increase in protection upon incorporation into the nucleosome. Enhanced protection of H4 appears to occur globally, with virtually the entire protein undergoing an increase in protection in the nucleosome compared to tetramer (Figure 4.14, difference plot and Figure 4.15). For H3, the core region of the protein experiences a dramatic increase in protection in the nucleosome compared to tetramer, while the tail domain experiences only a very small increase in protection in the first two time points (17% and 11% decrease in deuterium incorporation for the H3 tail in the nucleosome at 10^1s and 10^2s,

respectively) (Figure 4.15). Notably, the H4 tail domain undergoes a massive increase in protection, only reaching approximately 60% deuteration by 10^4 seconds of exchange within the nucleosome. This corresponds to a 36%-51% decrease in deuterium content of the tail region within the nucleosome compared to tetramer.



c

Histone H3				Histone H4			
Time Point	D	D-Crit	Significant?	Time Point	D	D-Crit	Significant?
10^1s	0.143	0.026	Yes	10^1s	0.162	0.033	Yes
10^2s	0.133	0.024	Yes	10^2s	0.122	0.030	Yes
10^3s	0.127	0.023	Yes	10^3s	0.094	0.028	Yes
10^4s	0.130	0.022	Yes	10^4s	0.079	0.026	Yes

Figure 4.14. Histones H3 and H4 experience a dramatic increase in protection upon incorporation into the nucleosome. Exchange profiles are given for H4 (A) and H3 (B) in

tetramer and nucleosome contexts. The bars above the heat maps indicate peptides that were used in the analysis. For the first two heat maps in each panel, the color scheme indicates the degree of deuteration as shown in the legend to the right. For the difference plots (the last plot for each panel), deuteration % of the nucleosome was subtracted from the tetramer. Blue indicates that the nucleosome is more protected and red indicates that the tetramer is more protected from exchange as shown in the legend to the right. Proline resides are shown as white spaces because they do not contain a backbone amide. All other white spaces indicate gaps in sequence coverage. (C) K-S two-sample test results to determine if H3 and H4 have different exchange profiles in the tetramer compared to nucleosome at each time point. A cumulative distribution of deuterium content (%) per residue for H3 and H4 in tetramer and nucleosome contexts was calculated. The maximum difference between the tetramer and nucleosome cumulative distributions for H3 and H4 (D) were determined for each time point and compared to the calculated D-critical value to determine if the difference is significant.

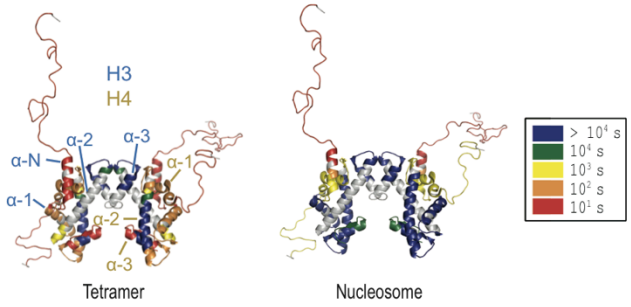


Figure 4.15. The enhanced protection of H4 in nucleosome context occurs globally while H3 experiences enhanced protection mainly in the core domain of the protein. Exchange profiles for H3 and H4 from the tetramer and nucleosome contexts are mapped to the crystal structure of the nucleosome (with DNA, H2A, and H2B hidden for easier viewing) (PDB: 1kx5). The color of each residue indicates the time point at which that residue was greater than 50%

deuterated. Gray color indicates that no information was available due to lack of sequence coverage for that region. Secondary structures of H3 (blue) and H4 (yellow) are indicated.

To determine whether the observed differences in exchange profiles for H3 and H4 in tetramer compared to nucleosome contexts are statistically significant, we performed a K-S two-sample test for each time point. The test compared the deuterium content of each residue in tetramer and nucleosome constructs for each protein. We found that all differences in exchange profiles between the tetramer and nucleosome are statistically significant for H3 and H4 (H3: 10s: D = 0.143, D-crit = 0.026; 100s: D = 0.133, D-crit = 0.024; 1,000s: D = 0.127, D-crit = 0.023; 10,000s: D = 0.130, D-crit = 0.022) (H4: 10s: D = 0.162, D-crit = 0.033; 100s: D = 0.122, D-crit = 0.030; 1,000s: D = 0.094, D-crit = 0.028; 10,000s: D = 0.079, D-crit = 0.026). These results indicate that H3 and H4 have different exchange profiles in the tetramer compared to nucleosomal context.

H2A and H2B were also analyzed in nucleosome context, and the results are displayed in Figure 4.14. There is a large degree of protection across the entire sequence for both proteins, with the exception of the H2B tail which experiences rapid exchange and is nearly fully deuterated by 10 seconds. The H2B tail ranges from 82% to 96% deuterium incorporation throughout the timecourse. The H2A tail, however, experiences a large degree of protection, with 51% deuterium incorporation at 10 seconds and reaching a maximum of 69% deuterium incorporation by 10^4 seconds (Figure 4.16).

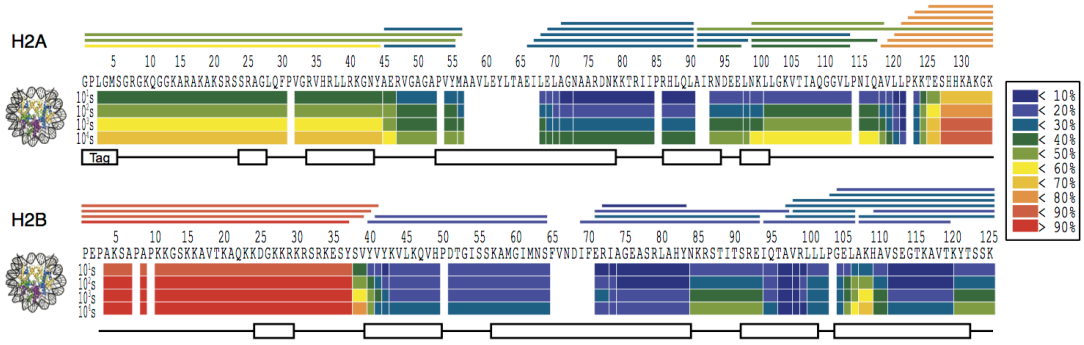


Figure 4.16. BU-HDX-MS analysis of H2A and H2B within the nucleosome show that areas of protection match with expected secondary structures. Exchange profiles are given for H2A and H2B in nucleosome context. The bars above the heat maps indicate peptides that were used in the analysis. The color scheme indicates the degree of deuteration as shown in the legend to the right. Values represent an average of three experimental replicates. Proline resides are shown as white spaces because they do not contain a backbone amide. All other white spaces indicate gaps in sequence coverage.

The regions of protection observed in these experiments align well with the secondary structures found in the crystal structure for all histones (Figure 4.15). However, the α-N helix of H3 is completely unstructured in tetramer context as indicated by extremely rapid deuterium exchange, but is protected in the nucleosome context, indicating that this helix may be protected by interaction with DNA and/or stabilized by the presence of H2A/H2B dimers within the nucleosome (Figure 4.14 and Figure 4.17). Globally, H3 and H4 in tetramer context are less protected from exchange compared to the nucleosomal context, indicating that incorporation into the nucleosome stabilizes H3 and H4 as has been observed previously (Black et al., 2004, 2007).

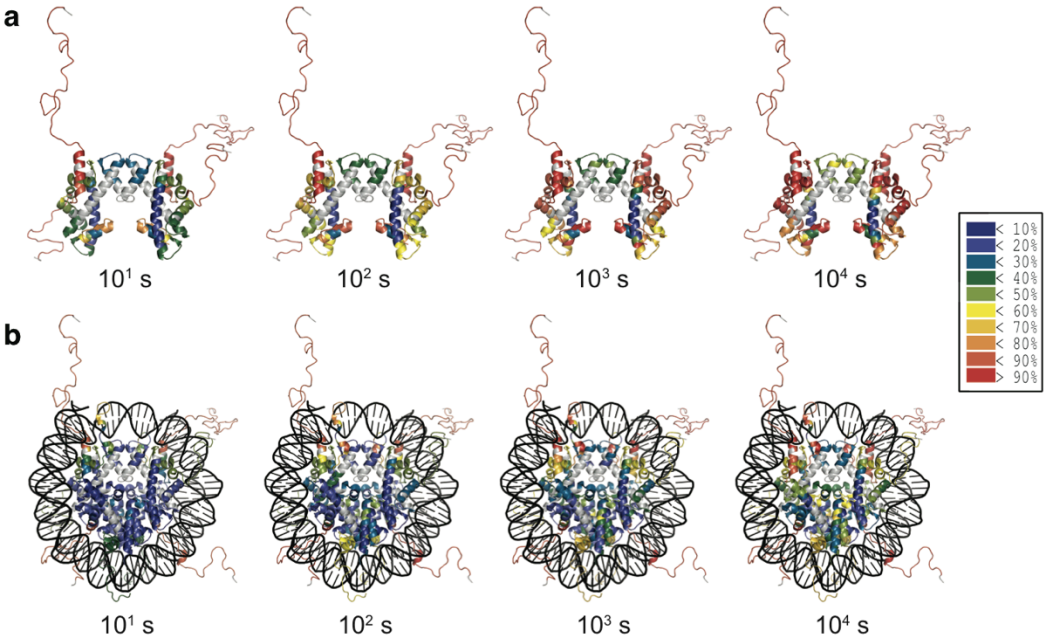


Figure 4.17. Global view of histone exchange data in tetramer and nucleosome. Exchange profiles for all histones in tetramer (A) and nucleosome (B) contexts are mapped to the crystal structure of the tetramer or nucleosome, respectively (PDB: 1kx5). The color of the residues indicates the % deuteration for that residue as indicated by the legend. Gray color represents gaps in sequence coverage where no information was obtained. Note that the tail domains were not solved in the crystal structure, but rather modeled in afterwards. The crystal structure contains histone sequences from *Xenopus laevis*, however the sequences are nearly identical to human (% identical to human: H4: 100%, H3: 99%, H2A: 95%, H2B: 93%).

The BU-HDX-MS data demonstrates that the H4 and H2A tails experience a large degree of protection within the nucleosome, while H3 and H2B tails have only a small degree of protection. However, the peptides containing the tail domains, ranging from 37 to 49 amino acids, are too long to localize areas of protection to specific regions within the tail. Therefore, we targeted a tail peptide for each histone for ETD fragmentation, allowing localization of deuterium content across the peptide in a MD-HDX-MS/MS experiment. We assessed the reproducibility of

deuterium content measurements for each fragment ion between experimental replicates (3 to 5 replicates for each time point). To this end, we calculated the CV of the deuterium content measurement for each detected fragment ion across all time points. We were able to achieve highly reproducible deuterium content measurements, with the median CV value being below 7% for each protein (Figure 4.18). H2A and H2B had the greatest degree of variability, likely because these tail peptides were lower in abundance than H3 and H4 and therefore had more variable isotopic distributions for the resulting fragment ions.

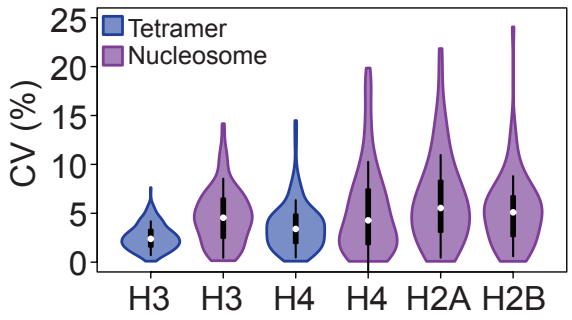


Figure 4.18. Deuterium content measurements for tail peptide fragment ions are highly reproducible in the MD-HDX-MS/MS platform. The CV values of the deuterium content of tail peptide fragment ions was calculated across all time points and plotted. White circles show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; polygons represent density estimates of data and extend to extreme values. Plot was generated using BoxPlotR (Spitzer et al., 2014).

Figure 4.19 displays the results of the MD-HDX-MS/MS experiment on histone tails. The results demonstrate that highly resolved deuterium content information was achieved for each

tail, reaching up to site-specific information depending on the quality of fragmentation. These results corroborate the findings in BU-HDX-MS experiment that H4 and H2A tails experience the greatest degree of protection, and H3 and H2B tails have a smaller degree of protection. However, we are now able to determine where these areas of protection exist. For H3, the small portion of the α-N helix is protected, as well as the region immediately adjacent to the predicted helix structure on the N-terminal side. For the H2B tail, there is protection in the first helix of the structure, as well as a small segment in the middle of the tail. H4 and H2A experience a more global protection, although the greatest protection is localized to the α-helix within the tail peptide.

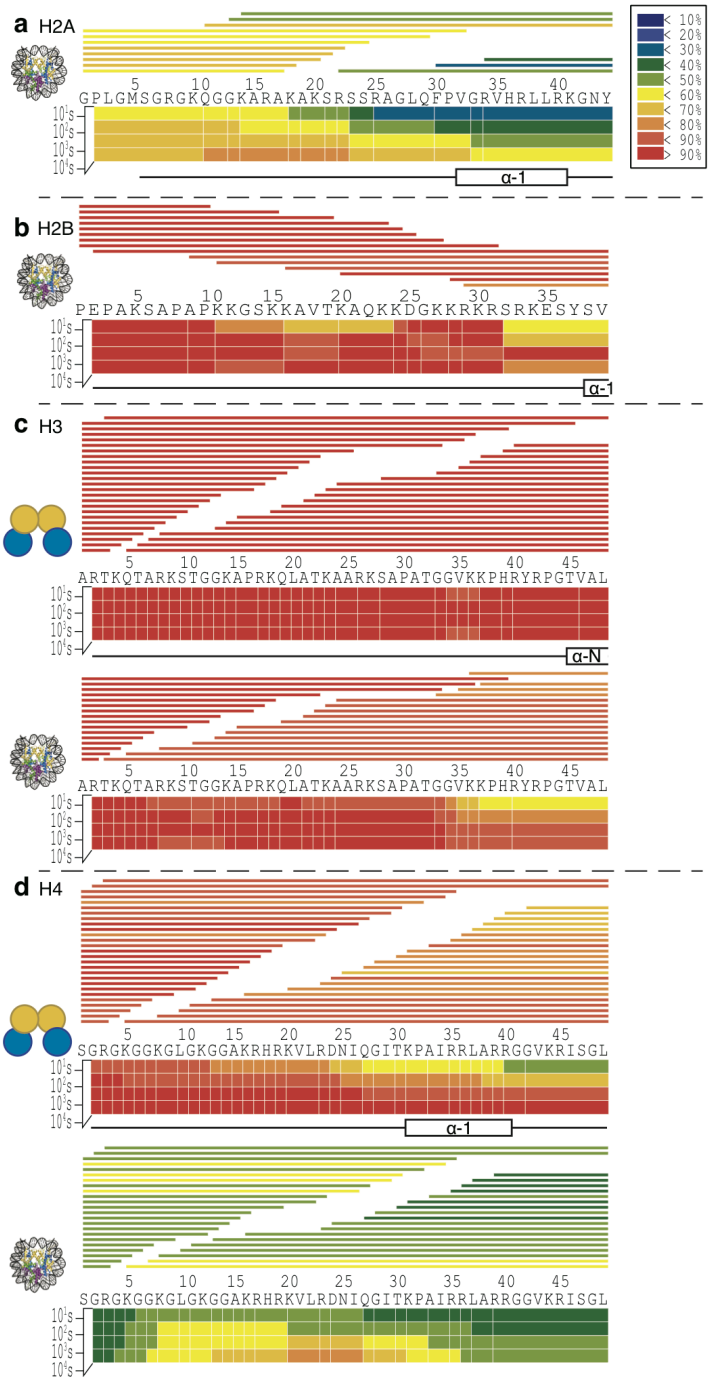
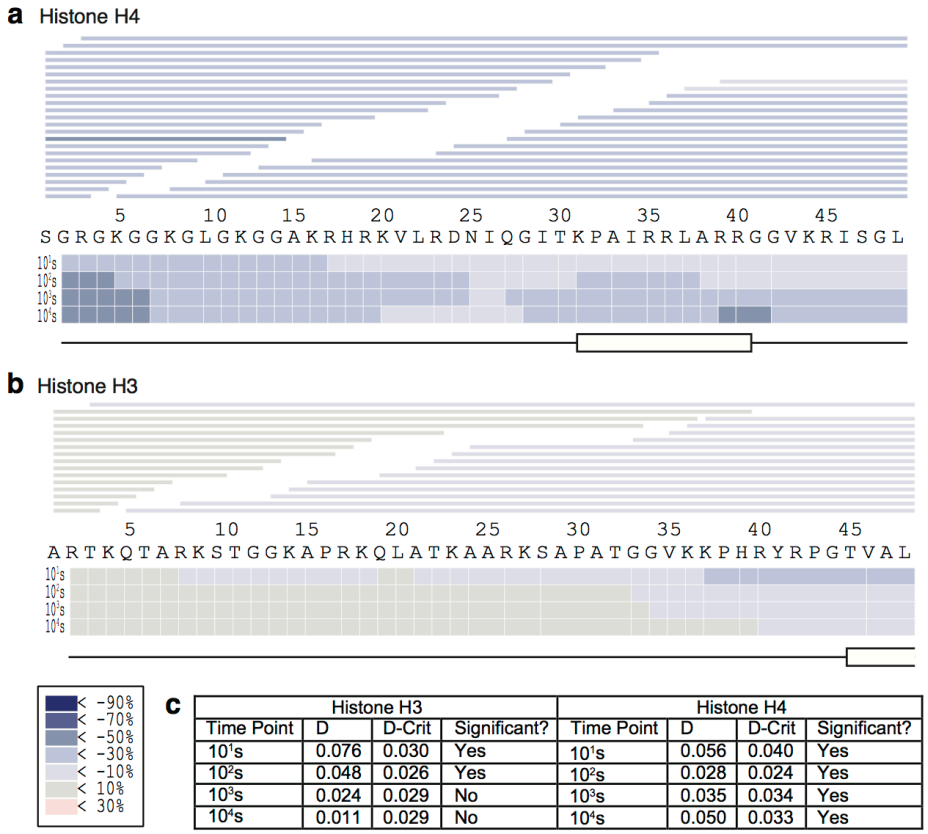


Figure 4.19. MD-HDX-MS/MS enables near site-specific resolution of deuterium content in tail domains of each histone. Exchange profiles are given for each histone tail: H2A (A), H2B (B), H3 (C), and H4 (D), in tetramer context (when relevant) and nucleosome context, as

indicated by the icon to the left. The bars above the heat maps indicate fragment ions that were used in the analysis. The color scheme indicates the degree of deuteration as shown in the legend to the right. Values represent an average of at least three experimental replicates. Secondary structures are mapped underneath each heat map.

The H3 and H4 tails exhibit an increased protection in the nucleosome relative to the tetramer (Figure 4.20). The degree of protection is much greater for H4, reaching differences in deuterium content of up to 64% between the tetramer and nucleosome constructs. However, H3 tails exhibit some increased protection within the nucleosome, mainly corresponding to the α-N helix, which reaches up to 44% greater protection in the nucleosome compared to the tetramer, which is nearly fully deuterated by 10 seconds. However, nearly the entire tail undergoes slight protection (around 10% lower deuterium content) in the 10 second time point, indicating that incorporation into the nucleosome confers a small degree of protection from exchange for the H3 tail relative to the tetramer, however the most dramatic difference is located in the α-N helix and region immediately adjacent. Together, these results show that incorporation into the nucleosome confers increased structural rigidity to the H3 and H4 tails, albeit to different degrees.



c

Histone H3				Histone H4			
Time Point	D	D-Crit	Significant?	Time Point	D	D-Crit	Significant?
10^1s	0.076	0.030	Yes	10^1s	0.056	0.040	Yes
10^2s	0.048	0.026	Yes	10^2s	0.028	0.024	Yes
10^3s	0.024	0.029	No	10^3s	0.035	0.034	Yes
10^4s	0.011	0.029	No	10^4s	0.050	0.033	Yes

Figure 4.20. H3 and H4 tail domains undergo enhanced protection upon incorporation into the nucleosome. Difference plots are given for (A) histone H4 tail and (B) histone H3 tail. The heat map demonstrates the difference in exchange between monomer and tetramer. Deuterium % of the tetramer was subtracted from the monomer, and the results are shown according to the legend. Blue indicates that the tetramer is more protected and red indicates that the monomer is more protected from exchange. (C) K-S two-sample test results to determine if H3 and H4 have different exchange profiles in the tetramer compared to nucleosome at each time point. A cumulative distribution of deuterium content (%) per residue for H3 and H4 in tetramer and nucleosome contexts was calculated. The maximum difference between the tetramer and nucleosome cumulative distributions for H3 and H4 (D) were determined for each time point and compared to the calculated D-critical value.

To determine whether the observed differences in exchange profiles for the H3 and H4 tails in the tetramer compared to nucleosome contexts are statistically significant, we performed a K-S two-sample test for each time point (Figure 4.20C). The results indicate that H4 has a significantly different exchange profile in tetramer compared to nucleosomal contexts across all time points. However, only the first two time points of the H3 exchange profile have a statistically significant difference, likely because the tail is nearly fully deuterated in both constructs by the last two time points.

4.2.3: Evaluation and comparison of BU-HDX-MS, MD-HDX-MS/MS, and TD-HDX-MS/MS platforms

As demonstrated in Chapter 4.2.2, the results from the BU, MD, and TD HDX platforms were in line with expected secondary structures according to the crystal structure. We compared the exchange profiles of H4 in tetramer context in each HDX-MS platform to determine how well the results agree (Figure 4.21A and B). To this end, we performed a Pearson correlation analysis for each time point based on the deuterium content of each amino acid. We found that each time point demonstrates a high degree of similarity between TD and MD platforms, with the following correlation coefficients: 10s: 0.837; 100s: 0.847; 1,000s: 0.842; 10,000s: 0.883. Each comparison was found to have statistically significant correlation values, with p-values less than 6×10^{-23}. We also performed a Pearson correlation analysis for all time points together and obtained a correlation coefficient of 0.858 (p-value: 1.71×10^{-29}), indicating that the TD- and MD-HDX-MS/MS platforms yielded highly similar results.

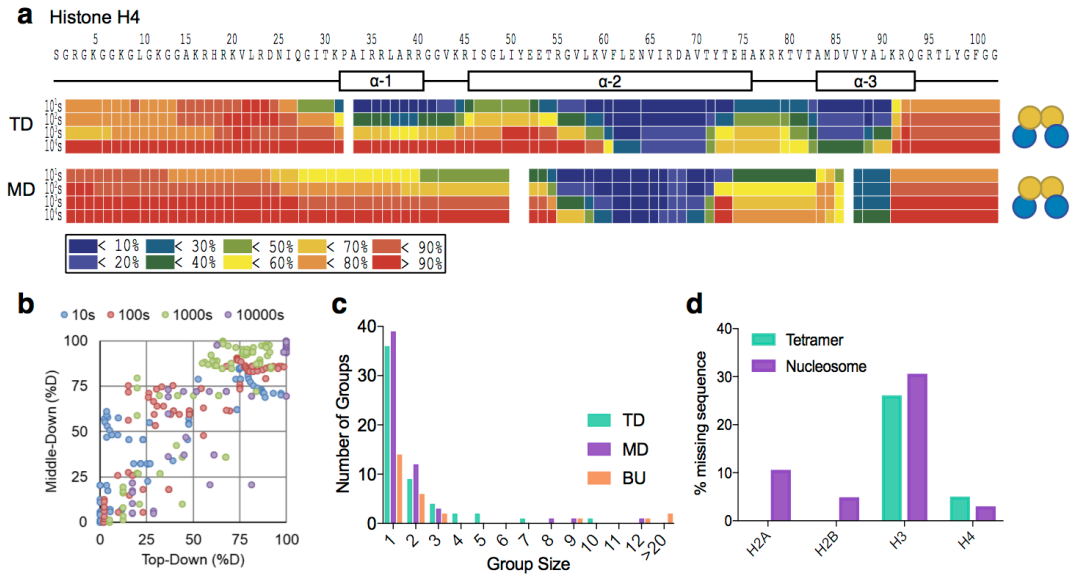


Figure 4.21. A comparison of the TD, BU, and MD HDX-MS platforms. (A) TD and MD provide comparable deuterium content measurements for histone H4 in tetramer context. Heat maps indicate the % deuteration of that residue according to the legend below the figure. The secondary structure is displayed above the heat maps. (B) Correlation plot for deuterium content of H4 in tetramer context for TD- and MD-HDX-MS/MS. Each point on the plot is color coded by time point and represents the deuterium content of a single residue. A Pearson correlation was performed for each time point with the following coefficients: 10s: 0.837; 100s: 0.847; 1,000s: 0.842; 10,000s: 0.882; all time points together: 0.858. The p-values were less than 6×10^{-27} for each correlation. (C) TD- and MD-HDX-MS/MS afford similar resolution of deuterium content localization for histone H4 in tetramer context. Group size indicates the number of residues that were analyzed together (i.e. 1 is site-specific information, 2 indicates that two amino acids were analyzed together, etc.). (D) MD-HDX-MS/MS results in sequence coverage gaps. The percent of amino acid sequence in which no information was obtained due to incomplete coverage of identified peptides is shown for each histone in nucleosome context and tetramer context (H3 and H4 only).

One of the greatest advantages of TD-HDX-MS/MS is that it has the potential to achieve site-resolved information across the entire protein sequence, depending on the quality of fragmentation. In MD-HDX-MS/MS, on the other hand, only certain peptides will be amenable to ETD fragmentation. In this case, only the tails had a high enough charge state and efficient fragmentation. Therefore, the other regions of the protein can only be studied at the intact level and will be dependent upon the presence of many overlapping peptides to get more site-resolved information, which is not always possible depending on the protein sequence. We analyzed the differences in resolution between the three platforms for H4 in tetramer form as this is the only protein that was analyzed across all three platforms (Figure 4.21C). To this end, we determined how many residues were analyzed together for each platform. The results demonstrate that the MD-HDX-MS/MS platform had a slightly higher number of singly resolved sites compared to TD-HDX-MS/MS; however, MD-HDX-MS/MS had a larger number of larger groups that had to be analyzed together (1 group of 8, 9, and 12 amino acids) compared to TD-HDX-MS/MS (1 group of 5, 7, and 10). As expected, BU-HDX-MS resulted in the worst level of resolution, with the lowest number of singly-resolved sites and more large groups (two of which are greater than 20 amino acids), highlighting the power of ETD-based HDX methods to improve the resolution of deuterium localization information compared to traditional BU-HDX-MS approaches.

Despite the fact that the resolution was not very different between MD-HDX-MS/MS and TD-HDX-MS/MS, the TD platform still has the advantage of guaranteeing full sequence coverage. The H4 protein had near complete coverage in MD-HDX-MS/MS; however, the other histone proteins had a greater amount of gaps in sequence coverage (Figure 4.21D). Therefore, while the resolution may be similar for H4, TD-HDX-MS/MS may be desirable for proteins in which MD-HDX-MS/MS does not yield full coverage.

4.3: Discussion

The results described here demonstrate the power of ETD-based HDX methods to obtain deuterium localization information in unprecedented detail. We were able to analyze, for the first time, a protein complex with TD-HDX-MS/MS and a heterogeneous protein complex and protein/DNA complex with MD-HDX-MS/MS. We were able to increase the number of site-specific deuterium localization from 14 sites in BU-HDX-MS to 36 and 39 sites in TD-HDX-MS/MS and MD-HDX-MS/MS, respectively. These HDX-MS/MS platforms can be tailored to target specific proteins or peptides of interest for ETD, demonstrating the adaptability and versatility of these methods.

Here, we use these ETD-based HDX platforms to provide the first detailed view of histone tail dynamics in different contexts, which has been an outstanding question in the field for decades. Previous work utilizing BU-HDX-MS techniques for histone analysis were able to provide insight into the structure and stability of core region of the protein but failed to provide detailed information about the tail domains. The Cleveland group demonstrated that H3 and H4 undergo massive increases in protection within the nucleosome compared to the tetramer (Black et al., 2004, 2007). Furthermore, D'Arcy et al. was able to demonstrate that the H2A/H2B dimer is largely unstable, undergoing nearly full exchange within 10 seconds under low salt conditions at room temperature (D'Arcy et al., 2013). While these studies provide critical insight into histone structure and dynamics, the employed methods are unable to characterize tail domains, either because they were not detected or because they were too long to get resolved information.

Our TD-HDX-MS/MS results demonstrate the capability of this platform to collect HDX information at unprecedented detail with high reproducibility. This method enables complete analysis of proteins as the digestion and separation steps present in MD and BU platforms is omitted, guaranteeing full coverage of the protein of interest. Our results also align with those of previous HDX and crystallography studies, demonstrating the reliability of the TD-HDX-MS/MS method. Observed areas of protection in H4 overlap nearly exactly with predicted secondary

structures observed in the crystal structure of the nucleosome. Furthermore, we found that the α-2 and α-3 helices of H4 exhibit a larger degree of protection than the α-1 helix, as was also observed in the HDX data collected by the Cleveland group (Black et al., 2004, 2007). We found that the H4 monomer also contains some areas of protection, mainly in the α-2 helix, indicating that the monomer may have some secondary structure, although it cannot be ruled out that some aggregation occurred during sample processing. One caveat of the TD-HDX-MS/MS approach is that direct infusion limits the complexity of samples that can be analyzed.

We demonstrated that MD-HDX-MS/MS, on the other hand, can accommodate complex samples due to the digestion and RP-HPLC separation steps through our analysis of tetramers and nucleosomes. We were able to achieve highly reproducible deuterium content measurements for the intact peptides as well as the fragment ions of the tail peptide (the median CV value was below 7% for each histone). We found that MD-HDX-MS/MS provides highly detailed information for ETD-amenable peptides, including the tail domains of the histone proteins, but can only provide resolution at the intact peptide level for the remainder of the protein. However, because many overlapping peptides were obtained in the digestion, we were able to match the resolution afforded by TD-HDX-MS/MS for histone H4 (Figure 4.21C). The other histones, however, had some sequence coverage gaps (ranging from 5-30% of the sequence) that preclude analysis of those regions. Analyzing these histones with TD-HDX-MS/MS (which was not completed here) would guarantee coverage of these regions.

Our MD-HDX-MS/MS results at the intact peptide level are corroborated by previous studies, indicating that this method is highly accurate and reliable. We found that H3 and H4 undergo a massive increase in protection upon incorporation into the nucleosome. This trend was observed for the core domains of H3 and H4 by the Cleveland group (Black et al., 2004, 2007). Furthermore, we found that the α-N helix of H3 appears to be highly unstructured in the tetramer but is protected from exchange in the nucleosomal context, indicating that this structure is

stabilized upon incorporation into the nucleosome, which is in line with previous studies (Black et al., 2007; Bowman et al., 2010).

The MD-HDX-MS/MS platform also allowed us to study histone tail domains in great detail for the first time. We found that the H3 tail is completely disordered in tetramer context, reaching near 100% deuteration by 10 seconds, including the region corresponding to the first half of the α-N helix (Figure 4.19). This was also observed in previous BU-HDX-MS data from Black and colleagues, where they found that a single H3 peptide spanning the length of the tail domain and half of the α-N helix was fully deuterated by 10 seconds at room temperature. However, we found that the H3 tail undergoes an increase in protection in the nucleosomal context compared to tetramer that was localized to the α-N helix and a 10 amino acid stretch immediately adjacent to the helix on the N-terminal side. This 10 amino acid section aligns with the portion of the tail that exits the nucleosomal core between the two DNA superhelical gyres close to the dyad axis, indicating that this protection may be due to interaction with the DNA (Figure 4.22A). Indeed, crosslinking studies indicate that the H3 tail can interact with DNA approximately 35-40 bp from the dyad (Zheng et al., 2005), precisely where the tail is mapped in the crystal structure (Luger et al., 1997).

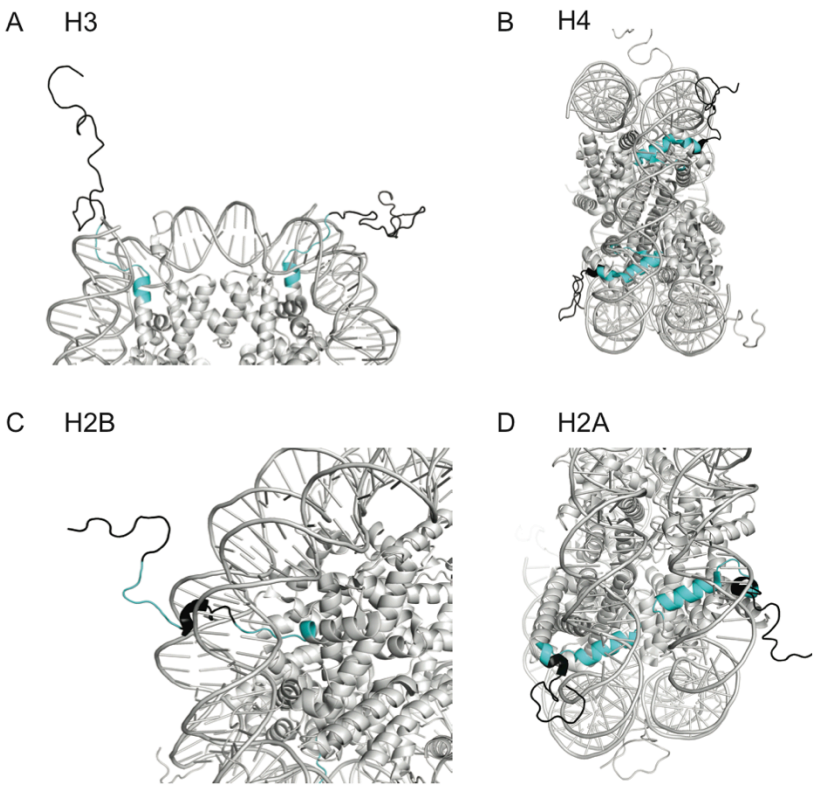


Figure 4.22. The histone tail domains are protected from exchange in nucleosomal context. Each tail peptide analyzed in MD-HDX-MS/MS is colored black, with the area of greatest protection at 10s are colored in cyan for H3 (A), H4 (B), H2B (C), and H2A (D). For H3, amino acids 35-48 are colored, which exhibit approximately 60% deuterium incorporation. For H4, residues 27-49, which exhibit approximately 40% deuterium incorporation, are colored. For H2B, residues 16-23 and 32-39 are colored, which exhibit approximately 60% deuterium incorporation. For H2A, residues 25-44 are colored, which exhibit less than 30% deuterium incorporation. PDB: 1kx5. Note that the tail domains in this structure were modeled in after solving the structure without the tail domains.

The H4 tail undergoes a dramatic increase in protection upon incorporation into the nucleosome relative to the tetramer. Within the tetramer, the H4 tail exhibits protection corresponding to the α-1 helix of the sequence, although this protection extends into the tail domain as well, albeit to a lesser degree (Figure 4.19D). Within the nucleosome, however, the tail exhibits a large degree of protection, only reaching between 40-80% deuteration levels by the last time point, indicating that the nucleosome context reduces the conformational flexibility of the tail. The H4 tail protrudes from the nucleosomal surface from under the DNA superhelical gyre (Figure 4.22B), placing it in a position where it could potentially interact with the nucleosomal DNA. Indeed, crosslinking studies have demonstrated that the H4 tail can interact with DNA approximately 25-40 bp from the dyad. However, previous studies have also shown that residues 16-24 of the histone tail can bind to a region on the H2A and H2B surface called the acidic patch, named for the 8 acidic residues located at this surface. Most of these studies have shown the H4 tail binding to the acidic patch of an adjacent nucleosome, an interaction which has been shown to be critical in chromatin compaction and folding into higher order structures (Arya and Schlick, 2006; Luger et al., 1997). However, Kan and colleagues also determined, using crosslinking strategies on mono-nucleosomes, that an H4 tail can bind the acidic patch within the same nucleosome (Kan et al., 2009). The dramatic protection from exchange within the nucleosome structure could therefore be explained by interaction with the acidic patch and/or DNA within the same nucleosome; however it is also possible that the H4 tail is binding to the acidic patch of other nucleosome molecules in solution. Further experiments are needed to determine which of these interactions is occurring in solution.

The H2A and H2B tail peptides were studied in nucleosomal context only. The H2B peptide exhibits some protection corresponding to the area containing the first α helix as expected but also contains an additional area of slight protection in the middle of the tail domain. This result indicates that there may be some secondary structure present or that this region interacts with the nucleosome. Indeed, the secondary structure prediction algorithm, PSIPRED, demonstrated that this region has a propensity for alpha-helix formation. Furthermore, the

modeled tails within the crystal structure do harbor a small α-helix in the tail domain, nearby the observed area of protection, although they do not overlap (Figure 4.22C).

The H2A tail peptide, on the other hand, experiences a large degree of protection from deuterium exchange in nucleosomal context, reaching a maximum of 80% deuteration levels by the last time point (10^4 seconds). This protection spans the entire length of the H2A tail as well as the first alpha-helix of the protein. The greatest protection in the tail domain is localized to the region immediately adjacent to the alpha-helix on the N-terminal side. The H2A tail exits the nucleosomal surface underneath the DNA superhelical gyre in a position where the tail could potentially interact with the nucleosomal DNA (Figure 4.22D). As seen in the crystal structure, where the tails are modeled in, H2A is found to interact with the DNA (Luger et al., 1997). This result is corroborated by the Hayes group who used crosslinking to show that residue A12 and G2 of the H2A tail can crosslink to the nucleosomal DNA 40 bp and 35-45 bp from the nucleosomal dyad, respectively (Lee and Hayes, 1997).

We also compared the performances of our BU-HDX-MS, MD-HDX-MS/MS, and TD-HDX-MS/MS for histone analysis. To this end, we compared the deuterium localization results of histone H4 in tetramer context, as this is the only protein that was analyzed across all three platforms (Figure 4.21). The results demonstrate that all three platforms generate highly similar results, although MD- and TD-HDX-MS/MS enable far superior resolution compared to BU-HDX-MS. For histone H4, TD- and MD-HDX-MS/MS yielded similar levels of resolution. This result is likely due to the fact that the pepsin digest for H4 yielded many overlapping peptides, enabling more site-specific deuterium localization to be obtained. However, full coverage for the other three histone proteins was not obtained in the MD-HDX-MS/MS platform, preventing analysis of these regions using this platform. TD-HDX-MS/MS ensures full coverage of the protein of interest and is therefore not a limitation in this approach.

Given that these methods returned highly similar results, determining the optimal platform for a protein of interest will depend upon the exact experimental conditions because each of the

platforms has different limitations and strengths. TD-HDX-MS/MS ensures full coverage of the protein, requires dramatically less instrument time, has fewer sample processing steps, and can achieve higher resolution than MD-HDX-MS/MS. However, because the digestion and separation steps are omitted in this platform, samples must be relatively simple to avoid spectral overlap. Furthermore, the protein of interest must be ETD-amenable or the resolution obtained will be poor. MD-HDX-MS/MS, on the other hand, can accommodate more complex samples and requires approximately 5- to 10-fold less protein. Additionally, the protein does not have to be ETD-amenable because any charge-poor regions of the protein can be analyzed at the intact peptide level.

Overall, these results demonstrate that MD-HDX-MS/MS and TD-HDX-MS/MS enable precise and highly reproducible deuterium localization of proteins in complex samples. This study represents the first heterogeneous protein complex and protein/DNA complex to be analyzed with ETD-based HDX-MS methodology, demonstrating the versatility and power of these methods for extremely detailed structural and dynamic analysis of protein molecules.

4.4: Methods

4.4.1: Protein expression, purification, and reconstitution

Human histones H2A, H2B, H3, and H4 were expressed in BL21 [DE3] (pLysS) cells and purified as monomers as previously described (Luger et al., 1999; Sekulic et al., 2010) . Briefly, protein was extracted from cells and separated by gel filtration (column: HP Sephacryl 26/60 S200). For H3, H4, and H2B, fractions containing the histone of interest were pooled, dialyzed into cation exchange buffer, and purified by cation exchange (column: HiTrap 5mL SPFF). Histone fractions were pooled, dialyzed into low salt buffer, and lyophilized for long-term storage. The histone sequences are shown in Table 4.2. Note that the expression vector contained an initial methionine residue that is cleaved off in the cells. H2A was engineered to contain a His-tag

and was purified using a nickel column followed by the same chromatographic procedures described above. The His-tag was cleaved with Precission protease to leave a small amino acid tag at the N-terminus with the sequence GPLG.

Table 4.2. Amino acid sequences of purified histone proteins.

Protein	Full protein sequence
H3	ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQK STELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEACEAYLVGLFEDTNLCAIHAK RVTIMPKDIQLARRIRGEA
H4	SGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGV LKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG
H2A	GPLGMSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPV YMAAVLEYLTAEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTIAQGGVL PNIQAVLLPKKTESHHKAKGK
H2B	PEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISS KAMGIMNSFVNDIFERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGT KAVTKYTSS

Histones were refolded into H2A/H2B dimers and $(H3/H4)_2$ tetramers by mixing equimolar amounts of the constituent proteins followed by dialysis as previously described (Luger et al., 1997; Sekulic et al., 2010). Briefly, histone proteins were resuspended in urea buffer. Urea was slowly dialyzed out to allow for histone partners to fold together, and the resulting complex was purified using gel filtration. Histone complexes were concentrated to approximately 1ug/uL.

DNA, consisting of the 195 base pair 601 positioning sequence, was made using PCR from a plasmid containing the sequence of interest. The PCR reactions were pooled, DNA was precipitated with ethanol and purified by cation exchange chromatography. Fractions containing the DNA of interest were pooled, precipitated and resuspended to an appropriate concentration for use in nucleosome reconstitution.

Nucleosomes were reconstituted as previously described by combining the components with the appropriate 1:2:1 molar ratio of DNA, H2A/H2B dimers, and $(H3/H4)_2$ tetramers. Slow

dialysis into low salt buffer allowed for formation of nucleosome particles, which were validated by native gel electrophoresis.

4.4.2: Top-Down HDX/MS

4.4.2.1: Scrambling analysis

We used peptide probe 'P1' to determine the degree of scrambling (AnaSpec, Inc.) as described (Rand and Jørgensen, 2007). The powedered peptide was dissolved in 99.9% D_2O (Sigma) to a concentration of 1 ug/uL and incubated at room temperature for at least 24 hours. Peptides were diluted 50-fold into exchange buffer (50% acetonitrile, 0.1% formic acid, pH 2.5 in H_2O) for 10 seconds followed by freezing on dry ice. Samples were thawed and infused into the mass spectrometer as described below and fragmented with ETD. Data was analyzed as previously described (Zehl et al., 2008), except that centroids were calculated using ExMS2 software from the Englander lab (Kan et al., 2011).

4.4.2.2: HDX and sample preparation

Lyophilized monomeric H4 proteins were resuspended in buffer (10mM Tris, 0.3mM EDTA in H_2O, pH 7.0) to a final concentration of approximately 1ug/uL. Tetramer was also analyzed at this concentration in the same buffer. Deuterium exchange was conducted by mixing 20uL of protein (20ug) with 60uL of on-exchange buffer (10mM Tris, 0.3mM EDTA in D_2O, pD 7.51) for the indicated time at 4˚C. The exchange reaction was quenched by the addition of 120uL quench buffer (0.8% formic acid). Samples were immediately desalted using home-made C8 stage tip columns as previously described, except that the wash buffer consisted of 0.8% formic acid in dH_2O, pH 2.25 and the elution buffer consisted of 75% acetonitrile/25% wash buffer (Lin and Garcia, 2012; Sidoli et al., 2016). All steps in the desalting process were precisely timed to

prevent differences in back exchange between samples. Before addition of the sample, the spin columns were activated with 75uL of methanol and washed with 75uL of quench buffer. The desalting steps are as follows: 1. Add sample, spin at 7000xg for 2 minutes; 2. Add 75uL of quench, spin at 7000xg for 50 seconds; 3. Switch collection tube to a new clean tube, 4. Add 20uL elution buffer, spin at 2400xg for 1 minute. Samples were transferred to a new pre-cooled tube and immediately flash frozen in liquid nitrogen.

4.4.2.3: Infusion into MS

Samples were thawed in an ice water bath and immediately loaded into the pre-cooled (4˚C) sample tray of an Advion Triversa Nanomate. The Nanomate picked up 8uL of sample followed by 2uL of air to prevent the sample from falling out of the tip. The sample was delivered to a chip containing electrospray nozzles and infused. The spray quality was optimized in the first 30 seconds of infusion by altering the spray voltage (1.6 to 2.2 V) or air pressure of the Nanomate (0.4 to 0.5 psi) if needed. To prevent back exchange, a cooling device was used to lower the temperature of the sample as it infused into the mass spectrometer as described by the Jørgensen group (Amon et al., 2012). Briefly, we coiled approximately 60 feet of copper tubing inside an insulated box. The box was filled with dry ice, and nitrogen gas was flowed through the tubing at 25 psi and aimed directly on top of the infusion tip carrying the sample. The position of the gas nozzle was aligned perpendicular to the spray tip to minimize interference with sample spray. The temperature of the nitrogen gas exiting the cooling apparatus reached approximately -10˚C.

4.4.2.4: Instrument Method

The capillary temperature was set to 150˚C. All data were acquired manually. A full MS scan was acquired in the Orbitrap for approximately 30 seconds (resolution: 60,000; Scan range: 500-1300 m/z; S-lens RF level: 60%; AGC target: 5.0e5; maximum injection time: 100ms; source fragmentation: 35eV; 1 microscan). During this time, the centroid of the intact H4 in +15 charge state was approximated by eye. ETD MS/MS scans were then acquired in the Orbitrap (resolution: 120,000; scan range: 150-2,000 m/z; S-lens RF level: 60%; AGC target: 5.0e4; maximum injection time: 100ms; source fragmentation: 35 eV). The approximate centroid was specified for the precursor m/z. Ions were isolated in the quadrupole (isolation window: 8m/z) and fragmented with ETD (10ms) for approximately 3 minutes.

4.4.2.5: Data Analysis

HDExaminer (Sierra Analytics, version 2.5) was used to analyze deuterium content of the fragment ions, allowing +1 to +15 charge states. The intact sequence of H4 (Table 4.2) was imported into the software. Raw files were uploaded, and data was analyzed from minutes 1 to 2 of the data acquisition. All identified fragment ions were manually validated.

The ETD HX module of ExMS2 was also used to process all raw files. The +15 charge state was analyzed and the intact sequence of H3 or H4 was specified for the parent fragment sequence. MS/MS spectra were averaged from 1-2 minutes, and the following settings were specified: m/z tolerance: 10 ppm; individual peak noise threshold: 300; summed peak noise threshold: 1000; fitting goodness threshold: 0.80. All peptides were validated in the “Manual Check/Clean” window. Interfering peaks from nearby isotopic distributions were removed, and any peptides with interfering isotopic distributions from overlapping ions were deleted from the analysis.

4.4.3: Middle-Down HDX-MS/MS

4.4.3.1: HDX and sample preparation

Deuterium exchange reactions were carried out on ice by mixing 5uL of protein (~5ug) with 15uL of on-exchange buffer (10mM Tris, 0.3mM EDTA in D_2O, pD 7.51). pH of the on-exchange buffer was adjusted using DCl. Reactions were quenched at the indicated time points by addition of 30 uL quench buffer (2.5M gndHCl, 0.8% formic acid, 10% glycerol in dH_2O). Samples were immediately flash frozen in liquid nitrogen and stored at -80˚C until MS analysis.

4.4.3.2: LC-MS/MS

Samples were thawed in ice water and injected into a cooled online sample processing system (3-6˚C) composed of a pepsin column, a C18 trap column, and a C18 analytical column. A Shimadzu LC-10AD pump was used to pump the sample through an immobilized pepsin column at 0.05 mL/min onto a C18 trap column (1x 5mm, C18 PepMap100, Thermo Scientific, P/N 160434). Pepsin was immobilized by coupling to Poros 20 AL support (Applied Biosystems) and packed into column housings of 2 mm x 2 cm (64 ml) (Upchurch). Peptides were eluted onto and separated by an analytical C18 column (50 x 0.3mm, Targa 3um C18 resin, Higgins Analytical, Serial No. 269232) by a reverse-phase gradient delivered by an Agilent 1100 Series pump at 6uL/min (Buffer A: 0.1% formic acid, 0.05% TFA in H_2O, pH2.25 at room temperature; Buffer B: 0.1% formic acid in acetonitrile). The gradient consists of the following steps: 10-55% B in 15 minutes, 55-95% B in 5 minutes. The column was washed with 95%B for 30 minutes followed by re-equilibration at 10% B for 10 minutes between runs. Long washes are required to prevent carry-over from previous runs.

The sample was sprayed into a Thermo Orbitrap Fusion mass spectrometer. The capillary temperature was set to 215˚C. For each construct, at least three non-deuterated samples were analyzed using HCD and ETD to identify peptides for data analysis. For the HCD method, a full MS scan was collected in the Orbitrap (resolution: 60,000; 360-1000 m/z, AGC target: $5x10^5$, maximum injection time: 50ms, RF lens: 55%) followed by a series of MS/MS scans for two seconds where ions are chosen for fragmentation sequentially based on their abundance. Fragment ions were measured in the ion trap (HCD collision energy: 30%, stepped collision energy: 5%, scan rate: rapid, maximum injection time: 200ms, AGC target: 1×10^4, centroid mode). For the ETD method, a full MS scan was collected in the Orbitrap (same settings as HCD method), followed by ETD MS/MS scans based on abundance for two seconds in the Orbitrap (charge states: 5-10, ETD reaction time: 20 ms, ETD reagent target: 2×10^5, resolution: 60,000, maximum injection time: 400ms, AGC target: 2×10^5, 3 microscans). Tail peptides were targeted for fragmentation during their elution times (sequences in Table 4.3, target masses and charges in Table 4.4). Each file was analyzed using Mascot in the Proteome Discoverer software to generate an exclusion list for the next non-deuterated run so that deeper coverage could be obtained.

Given that HCD fragmentation leads to scrambling, all deuterated samples were analyzed using only ETD fragmentation. The same ETD MS method was used as specified for the non-deuterated samples; however, different masses were targeted according to Table 4.4 during their respective elution times. In tetramer runs, only H3 and H4 tails were targeted. In nucleosome runs, all four tails were targeted.

Table 4.3. Amino acid sequences of tails analyzed in MD-HDX-MS/MS

Tail	Amino acid sequence
H3	ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVAL
H4	SGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGL
H2A	GPLGMSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNY
H2B	PEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRSRKESYSV

Table 4.4. Mass, charge, and retention time information for tails analyzed in MD-HDX-MS/MS

Tail	CS	RT (min)	Mass targeted for ETD					
			ND	10^1 s	10^2 s	10^3 s	10^4 s	FD
H3	+7	6-8.5	723.141	726.015	726.159	726.447	726.590	726.445
H4	+7	8.5-10	742.310	743.891	744.034	744.179	744.612	745.618
H2A	+7	8.5-11	671.241	672.677	672.965	673.397	673.542	674.690
H2B	+6	4.5-7	714.586	716.768	717.270	717.270	717.605	717.102

CS = charge state; ND = non-deuterated samples; FD = fully-deuterated samples

4.4.3.3: Data Analysis

For non-deuterated samples, peptides were identified using pFind 3.0 for HCD data (peptide tolerance: 10 ppm, MS/MS tolerance: 0.4 Da) and Mascot for ETD data (peptide tolerance: 10 ppm, MS/MS tolerance: 0.04 Da) using a database containing the recombinant histone sequences (Table 4.4). HDExaminer (Sierra Analytics, version 2.5) was used to analyze deuterium incorporation of peptides. For BU-HDX-MS/MS analysis, the full sequence of the histone protein was imported into the software (Table 4.2). The confident results of the pFind and Mascot searches were combined into a single CSV file and uploaded into the software as the "peptide pool." Raw files were uploaded and analyzed using only the charge states identified by the pFind and Mascot searches.

HDExaminer was also used to analyze the MD-HDX-MS/MS tail peptide fragment ions. The sequence of the tail peptide was imported into the software (Table 4.3). Raw files were uploaded in top-down analysis mode, and the retention time was specified according to Table 4.3. Charge states 1 to the charge of the peptide (i.e. +6 or +7) were considered for analysis.

4.4.3.4: Scrambling analysis

To determine the degree of scrambling in the MD-HDX-MS/MS platform, we compared the deuterium content of the H2A tail fragment ions to the theoretical deuterium content of these ions with 100% scrambling. Theoretical values were calculated as previously described (Masson et al., 2017). A Kolmogorov-Smirnov two-sample test was used to test whether the difference in deuterium profiles of the experimental and theoretical fragment ions was statistically significant.

CHAPTER 5: Conclusions

5.1: Summary

This dissertation highlights work I have done to improve mass spectrometry-based methodology to study the composition and dynamic properties of histone proteins. Mass spectrometry (MS) has emerged as a leading and powerful tool to identify and evaluate proteins. Among the many applications of MS-based experiments include identification and quantification of proteins and their post-translational modifications (PTMs), identification of novel PTMs, and evaluation of protein dynamics and stability when combined with hydrogen-deuterium exchange (HDX) methodology. In my dissertation work, these are the applications that I sought to improve.

In Chapter 2, we aimed to determine if low-resolution mass spectrometers are capable of robust analysis of global histone PTM profiles, given that these instruments are less expensive, easier to maintain, and more ubiquitous than high-resolution instruments (Karch et al., 2014). We first sought to determine if the traditional MS method, which is composed of a hybrid data independent acquisition (DDA) and selected reaction monitoring (SRM) method, can accurately identify and quantify histone PTMs on a low-resolution LTQ-Velos Pro (Thermo) using the high-resolution Orbitrap Velos Pro (Thermo) for comparison. In the DDA/SRM hybrid method, ions are selected for fragmentation in the sequential order of their abundance based on the full MS scan (DDA portion). Co-eluting isobaric peptides, or peptides with identical masses but different compositions, are targeted for fragmentation across their elution profiles (SRM portion) to enable separate quantification of these species on the MS/MS level. We found that the low-resolution LTQ Velos Pro cannot differentiate the nearly isobaric tri-methylation and acetylation marks (42.0470 Da and 42.0106 Da, respectively) based on mass alone while the high-resolution Orbitrap Velos Pro can. However, we demonstrate that these marks can be easily distinguished on the LTQ Velos Pro using relative retention time information or monitoring co-elution of the ambiguous peaks with heavy labeled synthetic peptides (Figure 2.7). Furthermore, we found that both instruments obtained highly reproducible relative abundance measurements. We also found

that both instruments produced highly similar abundance measurements (R^2 = 0.878 for H3 and R^2 = 0.834 for H4), and that correcting for differences in ionization efficiency between the instruments improves this correlation even further (R^2 = 0.579 for H3 9-17AA before correction, R^2 = 0.979 after correction) (Figure 2.6). Together, these results demonstrate that low-resolution mass spectrometers are capable robust histone PTM quantification, although high-resolution instruments are more easily suited for this task (Karch et al., 2014).

In Chapter 2, we also sought to determine if data independent acquisition (DIA) strategies are capable of robust histone PTM quantification compared to the traditional DDA/SRM hybrid approach and whether DIA methods can be conducted on low-resolution instruments (Sidoli et al., 2015). The DDA/SRM hybrid approach has been extremely powerful for histone PTM analysis; however, it does present several drawbacks: (1) variability associated with the somewhat stochastic nature of data-dependent sampling, (2) requirement of a priori knowledge about elution times of isobaric peptides and subsequent reliance on highly reproducible chromatography, (3) inability to re-mine data post-run, and (4) long duty cycle of the method, reducing the chances of choosing low-abundant ions for fragmentation. In DIA methods, a series of sequential MS/MS scans are acquired, each spanning a window of m/z values (typically ranging from 10-50m/z), which span the desired m/z range. Therefore, all ions within a window are fragmented together, ensuring that all ions present in the sample are fragmented across their entire elution range. This enables quantification of all co-eluting isobaric species present in the sample, not just the ones specified by the user. To determine if DIA methods are suitable and as reliable as DDA/SRM methods for histone PTM analysis, we compared a single histone sample on both platforms. We first optimized the DIA method to maximize the quality of histone PTM analysis for both high- and low-resolution detectors and then compared this optimized DIA method to the traditional DDA/SRM method. We found no differences in the DIA analysis on low- and high-resolution detectors, demonstrating that low-resolution instruments are fully capable of rigorous analysis of histone PTMs using DIA methodology. Furthermore, we found that the results of the DIA and DDA/SRM method are highly similar (correlation, R^2 = 0.97, slope = 0.99 for low-

resolution DIA; R^2 = 0.98, slope = 0.97 for high-resolution DIA) (Figure 2.10). However, the DIA methods allowed for quantification of a greater number of co-eluting isobaric species compared to the DDA/SRM method and also allowed for re-mining of data after the run was finished. Together, these results demonstrate the power and flexibility of DIA methodology for analysis of complex samples such as histones.

In Chapter 3, I highlight work I have done to identify and quantify ADP-ribosylation sites on histones during DNA damage (Karch et al., 2017). ADP-ribosylation is a post-translational modification that occurs on many proteins throughout the cell, and histone proteins are one of the main acceptors within the nucleus. ADP-ribosylation is catalyzed by ADP-ribose transferase (ART) enzymes, the most well-studied of which include PARP-1 and PARP-2. PARP-1 and PARP-2 are critical for the DNA damage response and can bind directly to DNA lesions, which causes a large increase in their catalytic activity (Langelier et al., 2012). However, the role of histone ADP-ribosylation in DNA damage detection and repair is largely unexplored. We sought to determine where these PTMs occur on histones as a first step towards this goal. To this end, we treated HeLa cells with dimethyl sulfate (DMS), a potent DNA damaging agent, over a time course. We extracted histones from the cells, enriched for the modification using boronate chromatography, derivatized the ADP-ribose modification to facilitate identification, and analyzed them by MS. We were able to identify 30 ADP-ribosylation sites on histones, 20 of which are novel (Table 3.2). Notably, the derivatization procedure enables identification of ADP-ribosylation sites on Asp/Glu residues, but not on Arg/Lys/Ser, which have been previously found to be modified with ADP-ribose (Altmeyer et al., 2009; Laing et al., 2011; Leidecker et al., 2016; McDonald and Moss, 1994; Rosenthal et al., 2015; Zhang et al., 2013). However, our results determined that Asp/Glu are the main acceptors of ADP-ribosylation on histones, indicating that many of the available sites are able to be discovered in this study, although more likely exist. Ten of the identified sites had a high enough abundance to be quantified across the treatment time course, and we found that each site increases in abundance over time (Figure 3.7C). We mapped the PTM locations to the crystal structure and found that nearly every exposed Asp/Glu residue

on the surface of the nucleosome was found to be modified, while there are several Asp/Glu residues in the interior of the nucleosome that were not found to be modified here or in any previous study (Figure 3.8). These results suggest that it is unlikely that modification of specific residues is important for DNA damage detection and repair, but rather they imply that ART enzymes act through a "brute force" mechanism in which extensive ADP-ribosylation of the nucleosomal surface enables relaxation of chromatin and subsequent repair of the DNA lesion.

Chapter 4 illustrates work I have done to expand existing hydrogen-deuterium exchange (HDX) methodology coupled to MS to allow for robust and detailed analysis of histone protein dynamics and stability. The traditional HDX-MS experiment employs bottom-up (BU) mass spectrometry whereby proteins are digested, separated by RP-HPLC, and monitored by MS. In this approach, peptides are not fragmented so resolution of deuterium content information is limited to the length of the peptide. Traditional BU-HDX-MS studies are not able to evaluate tail domains of histones in great detail because the peptides spanning this region are quite long-ranging from 25 to over 50 amino acids (Black et al., 2004, 2007; D'Arcy et al., 2013). However, studies have shown that electron transfer dissociation (ETD) allows for retention of deuterium labeling within the peptide in the gas phase and can therefore be used to get more resolved deuterium localization (Masson et al., 2017; Rand and Jørgensen, 2007; Rand et al., 2009; Zehl et al., 2008). We therefore sought to analyze histone proteins with HDX coupled to middle-down (MD) and top-down (TD) mass spectrometry. In MD, ETD-amenable peptides are targeted for fragmentation using the same experimental set-up as BU. In TD-HDX-MS/MS, however, digestion and separation steps are omitted and the intact protein is infused and fragmented with ETD. We analyzed histone H4 monomers and $(H3/H4)_2$ heterotetramers using the TD-HDX-MS/MS approach and were able to get highly resolved information across the length of the H4 protein in both constructs, including tail domains, representing the first TD-HDX-MS/MS study to analyze protein complexes. Areas of observed protection map to secondary structures and corroborate previous BU-HDX-MS results from other groups, highlighting the reliability and accuracy of this method (Black et al., 2004, 2007). We analyzed $(H3/H4)_2$ heterotetramers and intact

nucleosomes using our MD-HDX-MS/MS platform, representing the first heterogeneous protein complex and protein/DNA complex to be analyzed with this technology. We were able to select the tail of each of the four core histones for fragmentation and get highly resolved information for each, representing the first experimentally obtained detailed view of histone tail dynamics to date. We found that the H4 and H2A tails exhibit a large degree of protection from deuterium exchange in nucleosomal context, likely due to interaction with DNA or other histone proteins within the nucleosome. H3 and H2B tails also exhibit protection from exchange, albeit to a lesser degree. The areas of protection may be due to interaction with DNA or, in the case of H2B, formation of relatively unstable secondary structure. Together, the results of these studies highlight the versatility, reliability, and reproducibility of ETD-based HDX-MS/MS methodology. MD- and TD-HDX-MS/MS are some of the only existing methods to study protein structure and stability in solution in such fine detail. Furthermore, these platforms can accommodate a wide range of proteins and complexes, demonstrating that they are highly adaptable and versatile methods, particularly when HPLC separation is employed.

5.2: Future directions

In Chapter 2, we demonstrated the power of DIA MS methodology for analysis of complex histone samples. Histones are among the most post-translationally modified proteins in Eukaryotes and contain many isobaric PTMs, making them a particularly challenging protein to study. However, histone proteins are less complex than larger samples, such as full proteomes or metabolomes. Since the advent of DIA methodology in the early 2000's, scientists have expanded the capabilities of DIA methodology and demonstrated its ability to evaluate very complex and difficult sample types. Recently, DIA has rapidly gained popularity and has become a leading tool in MS-based quantification of samples. For example, DIA has been used to characterize the proteomes of yeast (Selevsek et al., 2015), C. elegans (Venable et al., 2004), and human tissues (Bruderer et al., 2015, 2016). DIA has also been used to define

phosphoproteomes (Parker et al., 2015), metabolomes, and in identification of biomarkers (Muntel et al., 2015; Ortea et al., 2016). Along with this boost in popularity of DIA methodology, software tools for analysis of DIA data have also been improving (Bilbao et al., 2015; Hu et al., 2016). DIA methodology has its own limitations, some of which are inherent to MS in general, including the time it takes to complete a scan which affects the length of the duty cycle, the dynamic range which affects how many ions can be identified, and the complexity of the MS/MS spectra which affects the ability to assign fragment ions to a specific precursor. Many of these limitations are inter-related. For example, if the scan speed is slow, the length of the duty cycle increases, forcing the user to use larger m/z windows to decrease the duty cycle. This, in turn, increases the complexity of the MS/MS spectra as more precursors are grouped together and decreases the chances that low-abundant signals will be able to be detected. Therefore, DIA methodology faces some of the same challenges as DDA/SRM hybrid methodology, namely that not all ions can be detected/identified and the limitations associated with dynamic range. Improvements in instrumentation will be needed to overcome these limitations. Improvements in scan speed without sacrificing resolution and increasing sensitivity will allow users to decrease the size of the m/z windows used (and consequently collect less complex spectra and reduce dynamic range issues) without increasing the duty cycle (and consequently maintaining a high standard of quantification). The speed, sensitivity, and resolution of mass spectrometers has vastly improved within the last five years, and we have not yet reached the physical limitations of hardware improvements (Hu et al., 2016; Richards et al., 2015). Another challenge with DIA methodology is the analysis and deconvolution of complex MS/MS spectra; however, many software tools exist and are continuing to be improved (Bruderer et al., 2015; Hu et al., 2016; Keller et al., 2015; MacLean et al., 2010; Röst et al., 2014; Tsou et al., 2015). As instrumentation and analysis tools continue to improve, so will the power and versatility of DIA methodology.

In Chapter 3, we were able to identify and quantify many histone ADP-ribosylation sites during DNA damage. We were able to identify 30 ADP-ribosylation sites, 20 of which are novel. However, because these experiments were done in vivo with the full complement of ART

enzymes available, it is not possible to determine which ARTs are responsible for this modification in this experiment. Indeed, it is likely that PARP-1 and PARP-2 are responsible for a majority of the sites, given that they can bind to the DNA and directly modify nearby proteins, including histones (Langelier et al., 2012). In vitro studies have also confirmed that histones are a substrate for these PARP enzymes (Messner et al., 2010). However, the use of ART-specific inhibitors during this process would allow us to determine which ART enzymes can modify histones in vivo, and whether there is any site specificity or redundancy between these enzymes. This information would shed insight into DNA damage detection and repair pathways mediated by ADP-ribosylation.

We hypothesized that ART enzymes ADP-ribosylate the nucleosomal surface, likely in a non-site specific manner. Given that nearly the entire surface of the nucleosome is capable of accepting ADP-ribose modifications, it seems very unlikely that specific sites are critical for repair processes. However, testing whether specific sites are important for this process is very challenging due to the high copy number of histone genes in mammalian cells. Recently, however, Rakhimova et al. was able to determine site specific roles for ADP-ribosylation on H2BE18 and H2BE19 during repair of DNA double strand breaks in a slime mold, *Dictyostelium*, that has a single copy of each histone gene (Rakhimova et al., 2017). Studies such as these will be important to determine site-specific functionality of histone ADP-ribosylation sites, but conducting these studies in mammals will be highly challenging.

We used dimethyl sulfate (DMS) to induce DNA damage in HeLa cells. DMS is an alkylating agent; it alkylates purine bases, preventing the proper hydrogen bonds from forming with the sister DNA strand (Suzuki et al., 1983). DNA repair occurs through the base excision repair (BER) pathway whereby the methylated base is removed and the nucleotide is subsequently excised, causing the transient formation of a single-strand DNA break (SSB). However, it has been shown that ART enzymes are involved in all DNA damage repair pathways, including homologous end joining (HR), non-homologous end-joining (NHEJ), and alternative-

NHEJ (alt-NHEJ) (Pears et al., 2012). These pathways have different mechanisms of repair and therefore may have different patterns of ADP-ribosylation. Future studies should aim to use different DNA damaging agents that induce these other repair pathways to determine if histone ADP-ribosylation patterns are altered relative to the BER pathway. Elucidation of acceptor ADP-ribosylation sites will increase our understanding of these vital processes.

Histone ADP-ribosylation has been implicated in several other nuclear processes in addition to its role in DNA damage repair, including transcription, DNA replication, and regulation of chromatin structure (Messner and Hottiger, 2011). Determination of histone ADP-ribosylation acceptor sites during these processes would help us elucidate their role in these processes as well.

In Chapter 4, we demonstrated that MD- and TD-HDX-MS/MS are able to obtain deuterium localization information in unprecedented detail and are thus extremely valuable techniques for studying a wide range of proteins and protein complexes. We applied these platforms to histone proteins and were able to achieve the first in-depth look at histone tail structure and stability in tetramer and nucleosome context. Therefore, these platforms, particularly MD-HDX-MS/MS which can handle complex samples, can be used to study histone biology and nuclear processes in greater detail. For example, we were able to determine that the H4 tail undergoes a dramatic increase in protection upon incorporation into the nucleosome, but we were unable to determine if this observation is due to interaction with DNA or the acidic patch of the nucleosome. Repeating the study with nucleosomes containing a mutated acidic patch that binds more or less tightly to the H4 tail can help elucidate these interactions. It has also been shown that PTMs, such as acetylation, can alter tail domain interaction and potentially structure. The effect of PTMs on nucleosome stability can be monitored with these ETD-based HDX-MS techniques. Furthermore, histone tails are critical for proper formation of higher order chromatin structure, however the interactions have not been precisely mapped, and the relative contribution of each tail remains unknown. Studying relaxed and condensed oligonucleosomal arrays can help

elucidate these interactions in detail. Furthermore, there are many histone mutations that are suspected to affect stability of the nucleosome, and these could be explored with the described HDX platforms. Additionally, many binding proteins, such as histone H1 and transcription factors can bind nucleosomes. These binding interactions can be mapped with HDX-MS/MS and the effects of this interaction of nucleosome stability can be monitored. The effect of different DNA sequences on nucleosome stability can also be probed.

5.3: The many hats of mass spectrometry

Mass spectrometry was invented over a century ago and was made possible by the physicist J.J. Thomson's revolutionary work on cathode rays. In the early 1910's, Thomson published work in which he was able to ionize atoms and molecules using electron impact ionization and measure their m/z values by observing their trajectories in electric or magnetic fields, and thus began the field of mass spectrometry. MS remained primarily a tool for physicists for the next 50 years until chemists began to utilize this technology in the oil industry to identify low molecular weight compounds in crude oil. These studies led to the advent of chemical ionization techniques to volatilize small organic compounds, allowing MS to be used for more applications in organic chemistry. At the same time, gas-phase fragmentation techniques were being developed, including collision induced dissociation (CID), which eventually led to the capability of determining structural details of organic compounds by the late 1960's. Matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) techniques were introduced in the 1980's (for which Koichi Tanaka and John Fenn received the Nobel Prize in Chemistry in 2002), enabling the ionization of large biomolecules, notably peptides and proteins. This technology allowed MS to become a leading tool in the cellular, molecular, and structural biology fields. Today, advances in separation techniques, tandem MS technology, bioinformatics tools, and instrumentation have all contributed to the rise of MS as a leading analytical tool in many scientific disciplines ranging from biology to geology to atmospheric science.

MS methodology is quickly becoming an indispensable tool in the biological sciences and the number of studies using these techniques continues to grow. Advances in MS instrumentation have made this technique more user-friendly, affordable, and versatile, enabling more researchers to incorporate MS into their research. Nowhere has this been more evident than in the proteomics field, which is arguably the most popular application of MS in the biological sciences. MS can define entire proteomes, including splice variants and PTMs, making it a revolutionary tool in many proteomics applications. Beyond defining proteomes, MS has an ever-growing list of applications in protein science including structural biology through the use of HDX or crosslinking, identification of interactomes of proteins, and identification of biomarkers for disease. Furthermore, improvements in MS instrumentation and methodology has allowed MS to be utilized in other branches of biological science, including the analysis of glycans, nucleic acids, metabolites, and lipids. MS remains one of the most versatile and enabling analytical tools in biological science, and as sample preparation techniques, instrumentation, and bioinformatics tools continue to advance, so too will the number and types of MS applications.

BIBLIOGRAPHY

Abzalimov, R.R., Kaplan, D.A., Easterling, M.L., and Kaltashov, I.A. (2009). Protein Conformations Can Be Probed in Top-Down HDX MS Experiments Utilizing Electron Transfer Dissociation of Protein Ions Without Hydrogen Scrambling. J. Am. Soc. Mass Spectrom. *20*, 1514–1517.

Abzalimov, R.R., Bobst, C.E., and Kaltashov, I.A. (2013). A New Approach to Measuring Protein Backbone Protection with High Spatial Resolution Using H/D Exchange and Electron Capture Dissociation. Anal. Chem. *85*, 9173–9180.

Adamietz, P., and Rudolph, A. (1984). ADP-ribosylation of nuclear proteins in vivo. Identification of histone H2B as a major acceptor for mono- and poly(ADP-ribose) in dimethyl sulfate-treated hepatoma AH 7974 cells. J. Biol. Chem. *259*, 6841–6846.

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198–207.

Afjehi-Sadat, L., and Garcia, B.A. (2013). Comprehending dynamic protein methylation with mass spectrometry. Curr. Opin. Chem. Biol. *17*, 12–19.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proc. Natl. Acad. Sci. U. S. A. *51*, 786.

Altmeyer, M., Messner, S., Hassa, P.O., Fey, M., and Hottiger, M.O. (2009). Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. Nucleic Acids Res. *37*, 3723–3738.

Amon, S., Trelle, M.B., Jensen, O.N., and Jørgensen, T.J.D. (2012). Spatially Resolved Protein Hydrogen Exchange Measured by Subzero-Cooled Chip-Based Nanoelectrospray Ionization Tandem Mass Spectrometry. Anal. Chem. *84*, 4467–4473.

Arents, G., and Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc. Natl. Acad. Sci. *92*, 11170–11174.

Arya, G., and Schlick, T. (2006). Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. Proc. Natl. Acad. Sci. U. S. A. *103*, 16236–16241.

Banères, J.-L., Martin, A., and Parello, J. (1997). The N tails of histones H3 and H4 adopt a highly structured conformation in the nucleosome. J. Mol. Biol. *273*, 503–508.

Banerjee, S., and Mazumdar, S. (2012). Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte.

Bang, E., Lee, C.-H., Yoon, J.-B., Lee, W., and Lee, D.W. (2001). Solution structures of the N-terminal domain of histone H4. Chem. Biol. Amp Drug Des. *58*, 389–398.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Bassett, E.A., DeNizio, J., Barnhart-Dailey, M.C., Panchenko, T., Sekulic, N., Rogers, D.J., Foltz, D.R., and Black, B.E. (2012). HJURP uses distinct CENP-A surfaces to recognize and to stabilize CENP-A/histone H4 for centromere assembly. Dev. Cell *22*, 749–762.

Bern, M., Finney, G., Hoopmann, M.R., Merrihew, G., Toth, M.J., and MacCoss, M.J. (2010). Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. Anal. Chem. *82*, 833–841.

Bilbao, A., Varesio, E., Luban, J., Strambio-De-Castillia, C., Hopfgartner, G., Müller, M., and Lisacek, F. (2015). Processing strategies and software solutions for data-independent acquisition in mass spectrometry. PROTEOMICS *15*, 964–980.

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., and Cleveland, D.W. (2004). Structural determinants for generating centromeric chromatin. Nature *430*, 578–582.

Black, B.E., Brock, M.A., Bédard, S., Woods, V.L., and Cleveland, D.W. (2007). An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. Proc. Natl. Acad. Sci. *104*, 5008–5013.

Boulikas, T. (1988). At least 60 ADP-ribosylated variant histones are present in nuclei from dimethylsulfate-treated and untreated cells. EMBO J. *7*, 57–67.

Bowman, A., Ward, R., El-Mkami, H., Owen-Hughes, T., and Norman, D.G. (2010). Probing the (H3-H4)2 histone tetramer structure using pulsed EPR spectroscopy combined with site-directed spin labelling. Nucleic Acids Res. *38*, 695–707.

Bredehorst, R., Wielckens, K., Gartemann, A., Lengyel, H., Klapproth, K., and Hilz, H. (1978). Two Different Types of Bonds Linking Single ADP-Ribose Residues Covalently to Proteins. Eur. J. Biochem. *92*, 129–135.

Britton, L.-M.P., Gonzales-Cope, M., Zee, B.M., and Garcia, B.A. (2011). Breaking the histone code with quantitative mass spectrometry. Expert Rev. Proteomics *8*, 631–643.

Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., et al. (2015). Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. Mol. Cell. Proteomics *14*, 1400–1410.

Bruderer, R., Bernhardt, O.M., Gandhi, T., and Reiter, L. (2016). High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. PROTEOMICS *16*, 2246–2256.

Cabelof, D.C., Raffoul, J.J., Yanamadala, S., Ganir, C., Guo, Z., and Heydari, A.R. (2002). Attenuation of DNA polymerase β-dependent base excision repair and increased DMS-induced mutagenicity in aged mice. Mutat. Res. *500*, 135.

Cao, X.-J., Zee, B.M., and Garcia, B.A. (2013). Heavy Methyl-SILAC Labeling Coupled with Liquid Chromatography and High-Resolution Mass Spectrometry to Study the Dynamics of Site-Specific Histone Methylation. In Gene Regulation, M. Bina, ed. (Humana Press), pp. 299–313.

Carrivain, P., Cournac, A., Lavelle, C., Lesne, A., Mozziconacci, J., Paillusson, F., Signon, L., Victor, J.-M., and Barbi, M. (2012). Electrostatics of DNA compaction in viruses, bacteria and eukaryotes: functional insights and evolutionary perspective. Soft Matter *8*, 9285–9301.

Dai, L., Peng, C., Montellier, E., Lu, Z., Chen, Y., Ishii, H., Debernardi, A., Buchou, T., Rousseaux, S., Jin, F., et al. (2014). Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. Nat. Chem. Biol. *10*, 365.

D'Amours, D., Desnoyers, S., D'Silva, I., and Poirier, G.G. (1999). Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. Biochem. J. *342 (Pt 2)*, 249–268.

Daniels, C.M., Ong, S.-E., and Leung, A.K.L. (2014). Phosphoproteomic approach to characterize protein mono- and poly(ADP-ribosyl)ation sites from cells. J. Proteome Res. *13*, 3510–3522.

Daniels, C.M., Thirawatananond, P., Ong, S.-E., Gabelli, S.B., and Leung, A.K.L. (2015). Nudix hydrolases degrade protein-conjugated ADP-ribose. Sci. Rep. *5*, srep18271.

D'Arcy, S., Martin, K.W., Panchenko, T., Chen, X., Bergeron, S., Stargell, L.A., Black, B.E., and Luger, K. (2013). Chaperone Nap1 shields histone surfaces used in a nucleosome and can put H2A-H2B in an unconventional tetrameric form. Mol. Cell *51*, 662–677.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J. Mol. Biol. *319*, 1097–1113.

DeNizio, J.E., Elsässer, S.J., and Black, B.E. (2014). DAXX co-folds with H3.3/H4 using high local stability conferred by the H3.3 variant recognition residues. Nucleic Acids Res. *42*, 4318–4331.

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J. Mol. Biol. *327*, 85–96.

Englander, S.W. (2006). Hydrogen exchange and mass spectrometry: A historical perspective. J. Am. Soc. Mass Spectrom. *17*, 1481–1489.

Eustermann, S., Wu, W.-F., Langelier, M.-F., Yang, J.-C., Easton, L.E., Riccio, A.A., Pascal, J.M., and Neuhaus, D. (2015). Structural Basis of Detection and Signaling of DNA Single-Strand Breaks by Human PARP-1. Mol. Cell *60*, 742–754.

Even-Faitelson, L., Hassan-Zadeh, V., Baghestani, Z., and Bazett-Jones, D.P. (2016). Coming to terms with chromatin structure. Chromosoma *125*, 95–110.

Falick, A.M., Hines, W.M., Medzihradszky, K.F., Baldwin, M.A., and Gibson, B.W. (1993). Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. J. Am. Soc. Mass Spectrom. *4*, 882–893.

Falk, S.J., Guo, L.Y., Sekulic, N., Smoak, E.M., Mani, T., Logsdon, G.A., Gupta, K., Jansen, L.E.T., Van Duyne, G.D., Vinogradov, S.A., et al. (2015). Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. Science *348*, 699–703.

Falk, S.J., Lee, J., Sekulic, N., Sennett, M.A., Lee, T.-H., and Black, B.E. (2016). CENP-C directs a structural transition of CENP-A nucleosomes mainly through sliding of DNA gyres. Nat. Struct. Mol. Biol. *23*, 204–208.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. Science *246*, 64–71.

Ferro, A.M., and Olivera, B.M. (1982). Poly(ADP-ribosylation) in vitro. Reaction parameters and enzyme mechanism. J. Biol. Chem. *257*, 7808–7813.

Fierz, B., and Muir, T.W. (2012). Chromatin as an expansive canvas for chemical biology. Nat. Chem. Biol. *8*, 417–427.

Fierz, B., Chatterjee, C., McGinty, R.K., Bar-Dagan, M., Raleigh, D.P., and Muir, T.W. (2011). Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. Nat. Chem. Biol. *7*, 113.

Fischle, W., Tseng, B.S., Dormann, H.L., Ueberheide, B.M., Garcia, B.A., Shabanowitz, J., Hunt, D.F., Funabiki, H., and Allis, C.D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Flanagan, T.W., and Brown, D.T. (2016). Molecular dynamics of histone H1. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1859*, 468–475.

Fuchs, S.M., Krajewski, K., Baker, R.W., Miller, V.L., and Strahl, B.D. (2011). Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition. Curr. Biol. CB *21*, 53–58.

Garcia, B.A., Mollah, S., Ueberheide, B.M., Busby, S.A., Muratore, T.L., Shabanowitz, J., and Hunt, D.F. (2007a). Chemical derivatization of histones for facilitated analysis by mass spectrometry. Nat. Protoc. *2*, 933–938.

Garcia, B.A., Pesavento, J.J., Mizzen, C.A., and Kelleher, N.L. (2007b). Pervasive combinatorial modification of histone H3 in human cells. Nat. Methods *4*, 487.

Gessner, C., Steinchen, W., Bédard, S., J. Skinner, J., Woods, V.L., Walsh, T.J., Bange, G., and Pantazatos, D.P. (2017). Computational method allowing Hydrogen-Deuterium Exchange Mass Spectrometry at single amide Resolution. Sci. Rep. *7*.

Gibson, B.A., and Kraus, W.L. (2012). New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. Nat. Rev. Mol. Cell Biol. *13*, 411–424.

Gilany, K., and Luc, M. (2010). Mass spectrometry-based proteomics in the life sciences: a review. J. Paramed. Sci. *1*.

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Mol. Cell. Proteomics *11*, O111.016717.

Going, C.C., Xia, Z., and Williams, E.R. (2016). Real-time HD Exchange Kinetics of Proteins from Buffered Aqueous Solution with Electrothermal Supercharging and Top-Down Tandem Mass Spectrometry. J. Am. Soc. Mass Spectrom. *27*, 1019–1027.

Gordon, F., Luger, K., and Hansen, J.C. (2005). The core histone N-terminal tail domains function independently and additively during salt-dependent oligomerization of nucleosomal arrays. J. Biol. Chem. *280*, 33701–33706.

Guo, L.Y., Allu, P.K., Zandarashvili, L., McKinley, K.L., Sekulic, N., Dawicki-McKenna, J.M., Fachinetti, D., Logsdon, G.A., Jamiolkowski, R.M., Cleveland, D.W., et al. (2017). Centromeres are maintained by fastening CENP-A to DNA and directing an arginine anchor-dependent nucleosome transition. Nat. Commun. *8*, 15775.

Hansen, J.C. (2002). CONFORMATIONAL DYNAMICS OF THE CHROMATIN FIBER IN SOLUTION: Determinants, Mechanisms, and Functions. Annu. Rev. Biophys. Biomol. Struct. *31*, 361.

Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: Contribution to chromatin structure and function. Gene *431*, 1–12.

Hassa, P.O., Haenni, S.S., Elser, M., and Hottiger, M.O. (2006). Nuclear ADP-Ribosylation Reactions in Mammalian Cells: Where Are We Today and Where Are We Going? Microbiol. Mol. Biol. Rev. *70*, 789–829.

Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B., and Ahmad, K. (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res. *19*, 460–469.

Hoerner, J.K., Xiao, H., and Kaltashov, I.A. (2005). Structural and dynamic characteristics of a partially folded state of ubiquitin revealed by hydrogen exchange mass spectrometry. Biochemistry (Mosc.) *44*, 11286–11294.

Hopfgartner, G., Tonoli, D., and Varesio, E. (2012). High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. Anal. Bioanal. Chem. *402*, 2587–2596.

Hu, A., Noble, W.S., and Wolf-Yadlin, A. (2016). Technical advances in proteomics: new developments in data-independent acquisition. F1000Research *5*, 419.

Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. J. Mass Spectrom. *40*, 430–443.

Huang, R.Y.-C., Garai, K., Frieden, C., and Gross, M.L. (2011). Hydrogen/Deuterium Exchange and Electron-Transfer Dissociation Mass Spectrometry Determine the Interface and Dynamics of Apolipoprotein E Oligomerization. Biochemistry (Mosc.) *50*, 9273–9282.

Huletsky, A., Niedergang, C., Fréchette, A., Aubin, R., Gaudreau, A., and Poirier, G.G. (1985). Sequential ADP-ribosylation pattern of nucleosomal histones. Eur. J. Biochem. *146*, 277–285.

Ishii, H., Kadonaga, J.T., and Ren, B. (2015). MPE-seq, a new method for the genome-wide analysis of chromatin structure. Proc. Natl. Acad. Sci. U. S. A. *112*, E3457-3465.

Jenuwein, T., and Allis, C.D. (2001). Translating the Histone Code. Science *293*, 1074–1080.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant–containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. Nat. Genet. *41*, 941.

Jørgensen, T.J.D., Bache, N., Roepstorff, P., Gårdsvoll, H., and Ploug, M. (2005a). Collisional Activation by MALDI Tandem Time-of-flight Mass Spectrometry Induces Intramolecular Migration of Amide Hydrogens in Protonated Peptides. Mol. Cell. Proteomics *4*, 1910–1919.

Jørgensen, T.J.D., Gårdsvoll, H., Ploug, M., and Roepstorff, P. (2005b). Intramolecular Migration of Amide Hydrogens in Protonated Peptides upon Collisional Activation. J. Am. Chem. Soc. *127*, 2785–2793.

Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K., and Hansen, J.C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. J. R. Soc. Interface *10*, 20121022.

Kaltashov, I.A., Bobst, C.E., and Abzalimov, R.R. (2009). H/D Exchange and Mass Spectrometry in the Studies of Protein Conformation and Dynamics: Is There a Need for a Top-Down Approach? Anal. Chem. *81*, 7892–7899.

Kan, P.-Y., Caterino, T.L., and Hayes, J.J. (2009). The H4 Tail Domain Participates in Intra- and Internucleosome Interactions with Protein and DNA during Folding and Oligomerization of Nucleosome Arrays. Mol. Cell. Biol. *29*, 538–546.

Kan, Z.-Y., Mayne, L., Chetty, P.S., and Englander, S.W. (2011). ExMS: Data Analysis for HX-MS Experiments. J. Am. Soc. Mass Spectrom. *22*, 1906–1915.

Kan, Z.-Y., Walters, B.T., Mayne, L., and Englander, S.W. (2013). Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. Proc. Natl. Acad. Sci. U. S. A. *110*, 16438–16443.

Karch, K.R., DeNizio, J.E., Black, B.E., and Garcia, B.A. (2013). Identification and interrogation of combinatorial histone modifications. Front. Epigenomics Epigenetics *4*, 264.

Karch, K.R., Zee, B.M., and Garcia, B.A. (2014). High resolution is not a strict requirement for characterization and quantification of histone post-translational modifications. J. Proteome Res.

Karch, K.R., Sidoli, S., and Garcia, B.A. (2016). Chapter One - Identification and Quantification of Histone PTMs Using High-Resolution Mass Spectrometry. In Methods in Enzymology, R. Marmorstein, ed. (Academic Press), pp. 3–29.

Karch, K.R., Langelier, M.-F., Pascal, J.M., and Garcia, B.A. (2017). The nucleosomal surface is the main target of histone ADP-ribosylation in response to DNA damage. Mol. Biosyst.

Kato, H., Van Ingen, H., Zhou, B.-R., Feng, H., Bustin, M., E Kay, L., and Bai, Y. (2011). Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. Proc. Natl. Acad. Sci. U. S. A. *108*, 12283–12288.

Keller, A., Bader, S.L., Shteynberg, D., Hood, L., and Moritz, R.L. (2015). Automated Validation of Results and Removal of Fragment Ion Interferences in Targeted Analysis of Data-independent Acquisition Mass Spectrometry (MS) using SWATHProphet. Mol. Cell. Proteomics MCP *14*, 1411–1418.

Kenzaki, H., and Takada, S. (2015). Partial Unwrapping and Histone Tail Dynamics in Nucleosome Revealed by Coarse-Grained Molecular Simulations. PLOS Comput. Biol. *11*, e1004443.

Kim, J.Y., Kim, K.W., Kwon, H.J., Lee, D.W., and Yoo, J.S. (2002). Probing Lysine Acetylation with a Modification-Specific Marker Ion Using High-Performance Liquid Chromatography/Electrospray-Mass Spectrometry with Collision-Induced Dissociation. Anal. Chem. *74*, 5443–5449.

Klammer, A.A., Yi, X., MacCoss, M.J., and Noble, W.S. (2007). Improving Tandem Mass Spectrum Identification Using Peptide Retention Time Prediction across Diverse Chromatography Conditions. Anal. Chem. *79*, 6111–6118.

Korolev, N., Yu, H., Lyubartsev, A.P., and Nordenskiöld, L. (2014). Molecular dynamics simulations demonstrate the regulation of DNA-DNA attraction by H4 histone tail acetylations and mutations. Biopolymers *101*, 1051–1064.

Krey, J.F., Wilmarth, P.A., Shin, J.-B., Klimek, J., Sherman, N.E., Jeffery, E.D., Choi, D., David, L.L., and Barr-Gillespie, P.G. (2013). Accurate Label-Free Protein Quantitation with High- and Low-Resolution Mass Spectrometers. J. Proteome Res.

Krishna, M.M.G., Hoang, L., Lin, Y., and Englander, S.W. (2004). Hydrogen exchange methods to study protein folding. Methods *34*, 51–64.

Krishnakumar, R., and Kraus, W.L. (2010). The PARP Side of the Nucleus: Molecular Actions, Physiological Outcomes, and Clinical Targets. Mol. Cell *39*, 8–24.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.-L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Mol. Cell *60*, 422–434.

Laing, S., Unger, M., Koch-Nolte, F., and Haag, F. (2011). ADP-ribosylation of arginine. Amino Acids *41*, 257–269.

Landgraf, R.R., Chalmers, M.J., and Griffin, P.R. (2012). Automated Hydrogen/Deuterium Exchange Electron Transfer Dissociation High Resolution Mass Spectrometry Measured at Single-Amide Resolution. J. Am. Soc. Mass Spectrom. *23*, 301–309.

Langelier, M.-F., Planck, J.L., Roy, S., and Pascal, J.M. (2012). Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. Science *336*, 728–732.

Langelier, M.-F., Steffen, J.D., Riccio, A.A., McCauley, M., and Pascal, J.M. (2017). Purification of DNA Damage-Dependent PARPs from E. coli for Structural and Biochemical Analysis. Methods Mol. Biol. Clifton NJ *1608*, 431–444.

Lee, K.-M., and Hayes, J.J. (1997). The N-terminal tail of histone H2A binds to two distinct sites within the nucleosome core. Proc. Natl. Acad. Sci. *94*, 8959–8964.

Lee, S., Tan, M., Dai, L., Kwon, O.K., Yang, J.S., Zhao, Y., and Chen, Y. (2013). MS/MS of Synthetic Peptide Is Not Sufficient to Confirm New Types of Protein Modifications. J. Proteome Res. *12*, 1007–1013.

Lehmann, K., Zhang, R., Schwarz, N., Gansen, A., Mücke, N., Langowski, J., and Toth, K. (2017). Effects of charge-modifying mutations in histone H2A α3-domain on nucleosome stability assessed by single-pair FRET and MD simulations. Sci. Rep. *7*, 13303.

Leidecker, O., Bonfiglio, J.J., Colby, T., Zhang, Q., Atanassov, I., Zaja, R., Palazzo, L., Stockum, A., Ahel, I., and Matic, I. (2016). Serine is a new target residue for endogenous ADP-ribosylation on histones. Nat. Chem. Biol. *12*, 998–1000.

Li, Z., and Kono, H. (2016). Distinct Roles of Histone H3 and H2A Tails in Nucleosome Stability. Sci. Rep. *6*, 31437.

Lin, S., and Garcia, B.A. (2012). Chapter One - Examining Histone Posttranslational Modification Patterns by High-Resolution Mass Spectrometry. In Methods in Enzymology, Carl Wu and C. David Allis, ed. (Academic Press), pp. 3–28.

Lin, S., Wein, S., Gonzales-Cope, M., Otte, G.L., Yuan, Z.-F., Afjehi-sadat, L., Maile, T., Berger, S.L., Rush, J., Lill, J.R., et al. (2014). Stable Isotope labeled histone peptide library for histone post-translational modification and variant quantification by mass spectrometry. Mol. Cell. Proteomics mcp.O113.036459.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.

Luger, K., Rechsteiner, T.J., and Richmond, T.J. (1999). Preparation of nucleosome core particle from recombinant histones. Methods Enzymol. *304*, 3–19.

Ma, D., Chan, M.K., Lockstone, H.E., Pietsch, S.R., Jones, D.N.C., Cilia, J., Hill, M.D., Robbins, M.J., Benzel, I.M., Umrania, Y., et al. (2009). Antipsychotic Treatment Alters Protein Expression Associated with Presynaptic Function and Nervous System Development in Rat Frontal Cortex. J. Proteome Res. *8*, 3284–3297.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968.

Makarov, A. (2000). Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Anal. Chem. *6*, 1156–1162.

March, R.E. (1997). An Introduction to Quadrupole Ion Trap Mass Spectrometry. J. Mass Spectrom. *32*, 351–369.

Martello, R., Leutert, M., Jungmichel, S., Bilan, V., Larsen, S.C., Young, C., Hottiger, M.O., and Nielsen, M.L. (2016). Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue. Nat. Commun. *7*, ncomms12917.

Martinez-Zamudio, R., and Ha, H.C. (2012). Histone ADP-Ribosylation Facilitates Gene Transcription by Directly Remodeling Nucleosomes. Mol. Cell. Biol. *32*, 2490–2502.

Masson, G.R., Maslen, S.L., and Williams, R.L. (2017). Analysis of phosphoinositide 3-kinase inhibitors by bottom-up electron-transfer dissociation hydrogen/deuterium exchange mass spectrometry. Biochem. J. *474*, 1867–1877.

Mayne, L., Kan, Z.-Y., Chetty, P.S., Ricciuti, A., Walters, B.T., and Englander, S.W. (2011). Many Overlapping Peptides for Protein Hydrogen Exchange Experiments by the Fragment Separation-Mass Spectrometry Method. J. Am. Soc. Mass Spectrom. *22*, 1898.

McDonald, L.J., and Moss, J. (1994). Enzymatic and nonenzymatic ADP-ribosylation of cysteine. Mol. Cell. Biochem. *138*, 221–226.

Messner, S., and Hottiger, M.O. (2011). Histone ADP-ribosylation in DNA repair, replication and transcription. Trends Cell Biol. *21*, 534–542.

Messner, S., Altmeyer, M., Zhao, H., Pozivil, A., Roschitzki, B., Gehrig, P., Rutishauser, D., Huang, D., Caflisch, A., and Hottiger, M.O. (2010). PARP1 ADP-ribosylates lysine residues of the core histone tails. Nucleic Acids Res. *38*, 6350–6362.

Mitchell Wells, J., and McLuckey, S.A. (2005). Collision-Induced Dissociation (CID) of Peptides and Proteins. In Methods in Enzymology, (Academic Press), pp. 148–185.

Moriwaki, Y., Yamane, T., Ohtomo, H., Ikeguchi, M., Kurita, J., Sato, M., Nagadoi, A., Shimojo, H., and Nishimura, Y. (2016). Solution structure of the isolated histone H2A-H2B heterodimer. Sci. Rep. *6*.

Mortusewicz, O., Amé, J.-C., Schreiber, V., and Leonhardt, H. (2007). Feedback-regulated poly(ADP-ribosyl)ation by PARP-1 is required for rapid response to DNA damage in living cells. Nucleic Acids Res. *35*, 7665–7675.

Moyle-Heyrman, G., Zaichuk, T., Xi, L., Zhang, Q., Uhlenbeck, O.C., Holmgren, R., Widom, J., and Wang, J.-P. (2013). Chemical map of Schizosaccharomyces pombe reveals species-specific features in nucleosome positioning. Proc. Natl. Acad. Sci. *110*, 20158–20163.

Muntel, J., Xuan, Y., Berger, S.T., Reiter, L., Bachur, R., Kentsis, A., and Steen, H. (2015). Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. J. Proteome Res. *14*, 4752–4762.

Murcia, G. de, Huletsky, A., Lamarre, D., Gaudreau, A., Pouyet, J., Daune, M., and Poirier, G.G. (1986). Modulation of chromatin superstructure induced by poly(ADP-ribose) synthesis and degradation. J. Biol. Chem. *261*, 7011–7017.

Nakaoka, S., Sasaki, K., Ito, A., Nakao, Y., and Yoshida, M. (2016). A Genetically Encoded FRET Probe to Detect Intranucleosomal Histone H3K9 or H3K14 Acetylation Using BRD4, a BET Family Member. ACS Chem. Biol. *11*, 729–733.

Niedergang, C.P., de MURCIA, G., It[illegible], M.-E., Pouyet, J., and Mandel, P. (1985). Time course of polynucleosome relaxation and ADP-ribosylation. Eur. J. Biochem. *146*, 185–191.

Ogata, N., Ueda, K., Kagamiyama, H., and Hayaishi, O. (1980). ADP-ribosylation of histone H1. Identification of glutamic acid residues 2, 14, and the COOH-terminal lysine residue as modification sites. J. Biol. Chem. *255*, 7616–7620.

Oka, J., Ueda, K., and Hayaishi, O. (1978). Snake venom phosphodiesterase: Simple purification with Blue Sepharose and its application to poly(ADP-ribose) study. Biochem. Biophys. Res. Commun. *80*, 841–848.

Ortea, I., Rodríguez-Ariza, A., Chicano-Gálvez, E., Arenas Vacas, M.S., and Jurado Gámez, B. (2016). Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. J. Proteomics *138*, 106–114.

Pan, J., and Borchers, C.H. (2014). Top-down mass spectrometry and hydrogen/deuterium exchange for comprehensive structural characterization of interferons: Implications for biosimilars. PROTEOMICS *14*, 1249–1258.

Pan, J., Han, J., Borchers, C.H., and Konermann, L. (2008). Electron Capture Dissociation of Electrosprayed Protein Ions for Spatially Resolved Hydrogen Exchange Measurements. J. Am. Chem. Soc. *130*, 11574–11575.

Pan, J., Han, J., Borchers, C.H., and Konermann, L. (2009). Hydrogen/Deuterium Exchange Mass Spectrometry with Top-Down Electron Capture Dissociation for Characterizing Structural Transitions of a 17 kDa Protein. J. Am. Chem. Soc. *131*, 12801–12808.

Pan, J., Han, J., Borchers, C.H., and Konermann, L. (2010). Characterizing Short-Lived Protein Folding Intermediates by Top-Down Hydrogen Exchange Mass Spectrometry. Anal. Chem. *82*, 8591–8597.

Pan, J., Han, J., Borchers, C.H., and Konermann, L. (2011). Conformer-Specific Hydrogen Exchange Analysis of Aβ(1–42) Oligomers by Top-Down Electron Capture Dissociation Mass Spectrometry. Anal. Chem. *83*, 5386–5393.

Pan, J., Han, J., Borchers, C.H., and Konermann, L. (2012). Structure and dynamics of small soluble Aβ(1-40) oligomers studied by top-down hydrogen exchange mass spectrometry. Biochemistry (Mosc.) *51*, 3694–3703.

Pan, J., Zhang, S., Parker, C.E., and Borchers, C.H. (2014). Subzero temperature chromatography and top-down mass spectrometry for protein higher-order structure characterization: method validation and application to therapeutic antibodies. J. Am. Chem. Soc. *136*, 13065–13071.

Pan, J., Zhang, S., Chou, A., Hardie, D.B., and Borchers, C.H. (2015). Fast Comparative Structural Characterization of Intact Therapeutic Antibodies Using Hydrogen–Deuterium Exchange and Electron Transfer Dissociation. Anal. Chem. *87*, 5884–5890.

Pan, J., Zhang, S., and Borchers, C.H. (2016). Comparative higher-order structure analysis of antibody biosimilars using combined bottom-up and top-down hydrogen-deuterium exchange mass spectrometry. Biochim. Biophys. Acta BBA - Proteins Proteomics *1864*, 1801–1808.

Panchenko, T., Sorensen, T.C., Woodcock, C.L., Kan, Z., Wood, S., Resch, M.G., Luger, K., Englander, S.W., Hansen, J.C., and Black, B.E. (2011). Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. Proc. Natl. Acad. Sci. *108*, 16588–16593.

Parker, B.L., Yang, G., Humphrey, S.J., Chaudhuri, R., Ma, X., Peterman, S., and James, D.E. (2015). Targeted phosphoproteomics of insulin signaling using data-independent acquisition mass spectrometry. Sci Signal *8*, rs6-rs6.

Patrie, S.M. (2016). Top-Down Mass Spectrometry: Proteomics to Proteoforms. SpringerLink 171–200.

Pears, C.J., Couto, C.A.-M., Wang, H.-Y., Borer, C., Kiely, R., and Lakin, N.D. (2012). The role of ADP-ribosylation in regulating DNA double-strand break repair. Cell Cycle Georget. Tex *11*, 48–56.

Peña, P.V., Musselman, C.A., Kuo, A.J., Gozani, and Kutateladze, T.G. (2009). NMR assignments and histone specificity of the ING2 PHD finger. Magn. Reson. Chem. MRC *47*, 352–358.

Perdigoto, C. (2017). Genome organization: Tracking chromosomal conformation through the cell cycle.

Pesavento, J.J., Bullock, C.R., LeDuc, R.D., Mizzen, C.A., and Kelleher, N.L. (2008). Combinatorial Modification of Human Histone H4 Quantitated by Two-dimensional Liquid Chromatography Coupled with Top Down Mass Spectrometry. J. Biol. Chem. *283*, 14927–14937.

Picotti, P., and Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat. Methods *9*, 555–566.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat. Biotechnol. *28*, 1057–1068.

Rakhimova, A., Ura, S., Hsu, D.-W., Wang, H.-Y., Pears, C.J., and Lakin, N.D. (2017). Site-specific ADP-ribosylation of histone H2B in response to DNA double strand breaks. Sci. Rep. *7*, 43750.

Rand, K.D., and Jørgensen, T.J.D. (2007). Development of a peptide probe for the occurrence of hydrogen (1H/2H) scrambling upon gas-phase fragmentation. Anal. Chem. *79*, 8686–8693.

Rand, K.D., Zehl, M., Jensen, O.N., and Jørgensen, T.J.D. (2009). Protein Hydrogen Exchange Measured at Single-Residue Resolution by Electron Transfer Dissociation Mass Spectrometry. Anal. Chem. *81*, 5577–5584.

Ren, C., Liu, S., Ghoshal, K., Hsu, P.-H., Jacob, S.T., Marcucci, G., and Freitas, M.A. (2007). Simultaneous metabolic labeling of cells with multiple amino acids: Localization and dynamics of histone acetylation and methylation. PROTEOMICS – Clin. Appl. *1*, 130–142.

Richards, A.L., Merrill, A.E., and Coon, J.J. (2015). Proteome sequencing goes deep. Curr. Opin. Chem. Biol. *0*, 11–17.

Riquelme, P.T., Burzio, L.O., and Koide, S.S. (1979). ADP ribosylation of rat liver lysine-rich histone in vitro. J. Biol. Chem. *254*, 3018–3028.

Rosenthal, F., Nanni, P., Barkow-Oesterreicher, S., and Hottiger, M.O. (2015). Optimization of LTQ-Orbitrap Mass Spectrometer Parameters for the Identification of ADP-Ribosylation Sites. J. Proteome Res. *14*, 4072–4079.

Röst, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmström, J., Malmström, L., et al. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat. Biotechnol. *32*, 219–223.

Rothbart, S.B., Dickson, B.M., Raab, J.R., Grzybowski, A.T., Krajewski, K., Guo, A.H., Shanle, E.K., Josefowicz, S.Z., Fuchs, S.M., Allis, C.D., et al. (2015). An Interactive Database for the Assessment of Histone Antibody Specificity. Mol. Cell *59*, 502–511.

Saikusa, K., Fuchigami, S., Takahashi, K., Asano, Y., Nagadoi, A., Tachiwana, H., Kurumizaka, H., Ikeguchi, M., Nishimura, Y., and Akashi, S. (2013). Gas-Phase Structure of the Histone Multimers Characterized by Ion Mobility Mass Spectrometry and Molecular Dynamics Simulation. Anal. Chem. *85*, 4165–4171.

Saikusa, K., Nagadoi, A., Hara, K., Fuchigami, S., Kurumizaka, H., Nishimura, Y., and Akashi, S. (2015). Mass spectrometric approach for characterizing the disordered tail regions of the histone H2A/H2B dimer. Anal. Chem. *87*, 2220–2227.

Sasaki, K., and Yoshida, M. (2014). Genetically encoded FRET indicators for live-cell imaging of histone acetylation. Methods Mol. Biol. Clifton NJ *1071*, 151–161.

Schneider, T.L., Halloran, K.T., Hillner, J.A., Conry, R.R., and Linton, B.R. (2013). Application of H/D Exchange to Hydrogen Bonding in Small Molecules. Chem. - Eur. J. *19*, 15101–15104.

Sekulic, N., Bassett, E.A., Rogers, D.J., and Black, B.E. (2010). The structure of (CENP-A–H4)$_2$ reveals physical features that mark centromeres. Nature *467*, 347.

Selevsek, N., Chang, C.-Y., Gillet, L.C., Navarro, P., Bernhardt, O.M., Reiter, L., Cheng, L.-Y., Vitek, O., and Aebersold, R. (2015). Reproducible and Consistent Quantification of the Saccharomyces cerevisiae Proteome by SWATH-mass spectrometry. Mol. Cell. Proteomics *14*, 739–749.

Sharma, A., Singh, K., and Almasan, A. (2012). Histone H2AX phosphorylation: a marker for DNA damage. Methods Mol. Biol. Clifton NJ *920*, 613–626.

Sidoli, S., Yuan, Z.-F., Lin, S., Karch, K., Wang, X., Bhanu, N., Arnaudo, A.M., Britton, L.-M., Cao, X.-J., Gonzales-Cope, M., et al. (2015a). Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis. Proteomics.

Sidoli, S., Lin, S., Xiong, L., Bhanu, N.V., Karch, K.R., Johansen, E., Hunter, C., Mollah, S., and Garcia, B.A. (2015b). Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH) Analysis for Characterization and Quantification of Histone Post-translational Modifications. Mol. Cell. Proteomics MCP *14*, 2420–2428.

Sidoli, S., Simithy, J., Karch, K.R., Kulej, K., and Garcia, B.A. (2015c). Low resolution data-independent acquisition in an LTQ-Orbitrap allows for simplified and fully untargeted analysis of histone modifications. Anal. Chem.

Sidoli, S., Bhanu, N.V., Karch, K.R., Wang, X., and Garcia, B.A. (2016). Complete Workflow for Analysis of Histone Post-translational Modifications Using Bottom-up Mass Spectrometry: From Histone Extraction to Data Analysis. J. Vis. Exp. JoVE.

Simithy, J., Sidoli, S., Yuan, Z.-F., Coradin, M., Bhanu, N.V., Marchione, D.M., Klein, B.J., Bazilevsky, G.A., McCullough, C.E., Magin, R.S., et al. (2017). Characterization of histone acylations links chromatin modifications with metabolism. Nat. Commun. *8*, 1141.

Smith, C.M., Gafken, P.R., Zhang, Z., Gottschling, D.E., Smith, J.B., and Smith, D.L. (2003). Mass spectrometric quantification of acetylation at specific lysines within the amino-terminal tail of histone H4. Anal. Biochem. *316*, 23–33.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P., and Li, G. (2014). Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. Science *344*, 376–380.

Spitzer, M., Wildenhain, J., Rappsilber, J., and Tyers, M. (2014). BoxPlotR: a web tool for generation of box plots. Nat. Methods *11*, 121.

Steen, H., and Mann, M. (2004). The abc's (and xyz's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. *5*, 699.

Steffen, J.D., McCauley, M.M., and Pascal, J.M. (2016). Fluorescent sensors of PARP-1 structural dynamics and allosteric regulation in response to DNA damage. Nucleic Acids Res. *44*, 9771–9783.

Sterling, H.J., and Williams, E.R. (2010). Real-time hydrogen/deuterium exchange kinetics via supercharged electrospray ionization tandem mass spectrometry. Anal. Chem. *82*, 9050–9057.

Strickfaden, H., McDonald, D., Kruhlak, M.J., Haince, J.-F., Th'ng, J.P.H., Rouleau, M., Ishibashi, T., Corry, G.N., Ausio, J., Underhill, D.A., et al. (2016). Poly(ADP-ribosyl)ation-dependent Transient Chromatin Decondensation and Histone Displacement following Laser Microirradiation. J. Biol. Chem. *291*, 1789–1802.

Suzuki, T., Kuwahara, J., and Sugiura, Y. (1983). Nucleotide sequence cleavage of guanine-modified DNA with aflatoxin B1, dimethyl sulfate, and mitomycin C by bleomycin and deoxyribonuclease I. Biochem. Biophys. Res. Commun. *117*, 916–922.

Switzar, L., Giera, M., and Niessen, W.M.A. (2013). Protein Digestion: An Overview of the Available Techniques and Recent Developments. J. Proteome Res. *12*, 1067–1077.

Syka, J.E.P., Marto, J.A., Bai, D.L., Horning, S., Senko, M.W., Schwartz, J.C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T., et al. (2004). Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-translational Modifications. J. Proteome Res. *3*, 621–626.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016.

Teif, V.B., and Bohinc, K. (2011). Condensed DNA: Condensing the concepts. Prog. Biophys. Mol. Biol. *105*, 208–222.

Thomas, C.E., Kelleher, N.L., and Mizzen, C.A. (2006). Mass Spectrometric Characterization of Human Histone H3: A Bird's Eye View. J. Proteome Res. *5*, 240–247.

Tse, C., and Hansen, J.C. (1997). Hybrid trypsinized nucleosomal arrays: identification of multiple functional roles of the H2A/H2B and H3/H4 N-termini in chromatin fiber compaction. Biochemistry (Mosc.) *36*, 11381–11388.

Tsou, C.-C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.-C., and Nesvizhskii, A.I. (2015). DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat. Methods *12*, 258–264, 7 p following 264.

Udeshi, N.D., Shabanowitz, J., Hunt, D.F., and Rose, K.L. (2007). Analysis of proteins and peptides on a chromatographic timescale by electron-transfer dissociation MS. FEBS J. *274*, 6269–6276.

Venable, J.D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., and Yates, J.R. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat. Methods *1*, 39–45.

Vera, D.L., Madzima, T.F., Labonne, J.D., Alam, M.P., Hoffman, G.G., Girimurugan, S.B., Zhang, J., McGinnis, K.M., Dennis, J.H., and Bass, H.W. (2014). Differential Nuclease Sensitivity Profiling of Chromatin Reveals Biochemical Footprints Coupled to Gene Expression and Functional DNA Elements in Maize. Plant Cell *26*, 3883–3893.

Verheugd, P., Bütepage, M., Eckei, L., and Lüscher, B. (2016). Players in ADP-ribosylation: Readers and Erasers. Curr. Protein Pept. Sci. *17*, 654–667.

Walters, B.T., Ricciuti, A., Mayne, L., and Englander, S.W. (2012). Minimizing Back Exchange in the Hydrogen Exchange-Mass Spectrometry Experiment. J. Am. Soc. Mass Spectrom. *23*, 2132–2139.

Walther, T.C., and Mann, M. (2010). Mass spectrometry–based proteomics in cell biology. J. Cell Biol. *190*, 491–500.

Wang, G., and Kaltashov, I.A. (2014). A new approach to characterization of the higher order structure of disulfide-containing proteins using hydrogen/deuterium exchange and top-down mass spectrometry. Anal. Chem.

Wang, G., Abzalimov, R.R., Bobst, C.E., and Kaltashov, I.A. (2013). Conformer-specific characterization of nonnative protein states using hydrogen exchange and top-down mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. *110*, 20087–20092.

Wang, L., Li, D.-Q., Fu, Y., Wang, H.-P., Zhang, J.-F., Yuan, Z.-F., Sun, R.-X., Zeng, R., He, S.-M., and Gao, W. (2007). pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. Rapid Commun. Mass Spectrom. *21*, 2985–2991.

Wang, X., Moore, S.C., Laszckzak, M., and Ausió, J. (2000). Acetylation Increases the α-Helical Content of the Histone Tails of the Nucleosome. J. Biol. Chem. *275*, 35013–35020.

Wei, H., and Yu, X. (2016). Functions of PARylation in DNA Damage Repair Pathways. Genomics Proteomics Bioinformatics *14*, 131–139.

Wiesner, J., Premsler, T., and Sickmann, A. (2008). Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. PROTEOMICS *8*, 4466–4483.

Wong, P.S.H., and Cooks, R.G. (1997). Ion Trap Mass Spectrometry. Curr. Sep. *16*, 85–92.

Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin higher-order structure and dynamics. Cold Spring Harb. Perspect. Biol. *2*, a000596.

Xiao, H., and Kaltashov, I.A. (2005). Transient Structural Disorder as a Facilitator of Protein-Ligand Binding: Native H/D Exchange—Mass Spectrometry Study of Cellular Retinoic Acid Binding Protein I. J. Am. Soc. Mass Spectrom. *16*, 869–879.

Yang, L., Tu, S., Ren, C., Bulloch, E.M.M., Liao, C.-L., Tsai, M.-D., and Freitas, M.A. (2010). Unambiguous determination of isobaric histone modifications by reversed-phase retention time and high-mass accuracy. Anal. Biochem. *396*, 13–22.

Young, N.L., DiMaggio, P.A., Plazas-Mayorca, M.D., Baliban, R.C., Floudas, C.A., and Garcia, B.A. (2009). High Throughput Characterization of Combinatorial Histone Codes. Mol. Cell. Proteomics *8*, 2266–2284.

Yuan, Z.-F., Lin, S., Molden, R.C., Cao, X.-J., Bhanu, N.V., Wang, X., Sidoli, S., Liu, S., and Garcia, B.A. (2015). EpiProfile Quantifies Histone Peptides With Modifications by Extracting Retention Time and Intensity in High-resolution Mass Spectra. Mol. Cell. Proteomics MCP *14*, 1696–1707.

Zee, B.M., Levin, R.S., Dimaggio, P.A., and Garcia, B.A. (2010). Global turnover of histone post-translational modifications and variants in human cells. Epigenetics Chromatin *3*, 22.

Zehl, M., Rand, K.D., Jensen, O.N., and Jørgensen, T.J.D. (2008). Electron Transfer Dissociation Facilitates the Measurement of Deuterium Incorporation into Selectively Labeled Peptides with Single Residue Resolution. J. Am. Chem. Soc. *130*, 17453–17459.

Zhang, K., Yau, P.M., Chandrasekhar, B., New, R., Kondrat, R., Imai, B.S., and Bradbury, M.E. (2004). Differentiation between peptides containing acetylated or tri-methylated lysines by mass spectrometry: An application for determining lysine 9 acetylation and methylation of histone H3. PROTEOMICS *4*, 1–10.

Zhang, Y., Wang, J., Ding, M., and Yu, Y. (2013). Site-specific characterization of the Asp- and Glu-ADP-ribosylated proteome. Nat. Methods *10*, 981–984.

Zhao, Y., and Garcia, B.A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. Cold Spring Harb. Perspect. Biol. *7*, a025064.

Zheng, Y., and Cui, Q. (2015). The histone H3 N-terminal tail: a computational analysis of the free energy landscape and kinetics. Phys. Chem. Chem. Phys. PCCP *17*, 13689–13698.

Zheng, C., Lu, X., Hansen, J.C., and Hayes, J.J. (2005). Salt-dependent Intra- and Internucleosomal Interactions of the H3 Tail Domain in a Model Oligonucleosomal Array. J. Biol. Chem. *280*, 33552–33557.

Zhou, B.-R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T.S., and Bai, Y. (2015). Structural Mechanisms of Nucleosome Recognition by Linker Histones. Mol. Cell *59*, 628–638.